

In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institution shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

1/2

SOME APPLICATIONS OF TOPOLOGY AND
FUNCTIONAL ANALYSIS IN PROBABILITY THEORY

A THESIS

Presented to

The Faculty of the Graduate Division

by

Charles McDonald Johnson, Jr.

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Applied Mathematics

Georgia Institute of Technology

August, 1964

Approved:

(Handwritten marks)

Date approved by Chairman: 12 Aug 1964

ACKNOWLEDGMENTS

The author gratefully acknowledges his debt to Professors James W. Walker, Eric R. Immel, and Robert H. Kasriel of the School of Mathematics, Georgia Institute of Technology, for their assistance and encouragement in the preparation of this thesis.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
CHAPTER	
I. INTRODUCTION	1
II. BASIC CONCEPTS	5
III. MEASURES AND INTEGRALS	24
IV. PROBABILITY THEORY	53
V. FUNCTIONAL ANALYSIS AND TOPOLOGY	71
VI. REGULAR MEASURES AND THE CONSISTENCY THEOREM	98
APPENDIX	116
LIST OF ABBREVIATIONS AND SYMBOLS	
BIBLIOGRAPHY	121

CHAPTER I

INTRODUCTION

The purpose of this paper is to study the foundations of the theory of generalized random variables of the form $X : \Omega \rightarrow X$, where Ω is a probability space and X is a topological measurable space. Many random phenomena in physics and engineering have outcomes which are not numbers, but vectors, matrices, functions, or operators. Thus it is desirable to extend the theory of probability to include such phenomena. A considerable amount of research has been devoted to this extension in recent years, for example in papers by SEGAL, BLACKWELL, DUBINS, HANS, NEDOMA, DRIML and HANS, SPACEK, and many others. A discussion of much of this recent work and an extensive bibliography can be found in the recently published book of GRENANDER. But the prerequisites for understanding the literature are formidable. Thus in this paper an attempt is made to present those parts of topology and functional analysis which will facilitate an understanding of the current literature.

Assumed to begin with are a knowledge of elementary set theory and logic, the properties of the real and complex number systems, and the basic theory of functions of a real variable. Chapter II is devoted to the presentation of the fundamental algebraic and topological concepts, and contains many of the definitions to be used throughout the remaining chapters.

In Chapter III the necessary concepts and tools from measure theory and integration are developed, first for real and complex valued

functions, then the theory is extended to Banach valued functions.

The theory of the Bochner integral presented here is a modified version of that found in HILLE and PHILLIPS; namely the domain measure space is assumed to be σ -finite, and simple functions are used in the definition of the Bochner integral instead of elementary functions. This makes it possible to develop much of the theory without using the Hahn-Banach Theorem, by using Egorov's Theorem and some of its consequences. The necessary parts of Hilbert space theory are examined briefly at the end of the chapter.

In Chapter IV the previously developed material is applied to extend some of the most important theorems of classical probability theory, from the real case to the Hilbert or Banach valued case. The material on conditional expectations is mostly due to DRIML and HANS.

In the first four chapters, no essentially topological concepts are used, except those which can be expressed in terms of sequential convergence in a metric space. Chapter V is devoted to the development of some of the functional analysis and topology which seems to be necessary in reading the current literature on generalized random variables, including the Hahn-Banach Theorem and the Tychonov Product Theorem and some of their consequences. The concept of compactness is particularly stressed here, since it is used extensively in Chapter VI.

In Chapter VI a generalization of the Kolmogorov Consistency Theorem is proved, using topological methods and a slight modification of Kolmogorov's proof as given in KOLMOGOROV [1], p. 29-33 and LOEVE, p. 93. See also HALMOS [1], p. 212. The Consistency Theorem has been generalized previously by SEGAL, BLACKWELL, and SPACEK; but these

results will not be examined here, since they were uncovered only after this paper had been almost completed. It will be noted only that the proof given here seems to be more direct and to use more elementary techniques, so that this version may be a special case of some of the others. The generalization proved here guarantees in particular the existence of random variables in a separable, complete pseudo-metric space which have preassigned, consistent joint distributions.

Concepts will be presented in an abstract, axiomatic form, usually without obvious intuitive motivation or description of applications.

There are several conventions with regard to notation and terminology which will be followed throughout. Sets will be denoted by capital letters C , and classes of sets by capital script letters ζ . The index n varies from 1 to ∞ , $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} a_n$, and $\bigcup A_n = \bigcup_{n=1}^{\infty} A_n$ unless otherwise indicated; similarly for m, k, j . The word sum, used in connection with sets, means disjoint union, and finite or countable sums of sets A_n will usually be denoted by ΣA_n or $A_1 + A_2 + \dots$ instead of $\bigcup A_n$; the use of this notation is intended to imply automatically that the sets A_n are disjoint. Unqualified statements involving a variable x in a space X are understood to mean "for all $x \in X$ " or "for arbitrary fixed $x \in X$ "; the meaning should always be clear from the context.

Names of references are capitalized, and multiple references by the same author are followed by numbers in square brackets. The references used most extensively are DRIML and HANS, HALMOS [1], [2], HILLE and PHILLIPS, KELLEY, KOLMOGOROV [1], [2], and LOEVE.

A complete list of symbols and abbreviations can be found in the Appendix.

In general, those theorems which are considered to be especially important, or which are generalizations of theorems in the literature, are given greater typographical emphasis.

CHAPTER II

BASIC CONCEPTS

In this chapter, the basic concepts to be used later will be introduced, and some of the relations between them will be examined.

These basic concepts are:

- (1) linear (vector) space and linear function
- (2) topological space and continuous function
- (3) measurable space and measurable function .

The algebraic concepts will be introduced first. A groupoid is a pair $(G, *)$ consisting of a nonempty set G and a binary operator $*: G \times G \rightarrow G$. A semigroup is an associative groupoid $(G, *)$: $(a * b) * c = a * (b * c)$. A monoid is a semigroup $(G, *)$ with an identity element $i \in G$ such that $i * a = a * i = a$ for all $a \in G$; an element $e \in G$ is a unit if there is an inverse element $e^{-1} \in G$ such that $e^{-1} * e = e * e^{-1} = i$. A group is a monoid $(G, *)$ such that every element in G is a unit; and an Abelian group is a commutative group: $a * b = b * a$. A ring is a triple $(R, +, *)$ such that $(R, +)$ is an (additive) Abelian group with (additive) identity θ , $(R, *)$ is a semigroup, and the distributive laws hold: $a * (b + c) = a * b + a * c$, $(a + b) * c = a * c + b * c$. A field (algebraic) is a ring $(F, +, *)$, with at least two elements, such that $(F - [0], *)$ is an Abelian group with identity 1, where 0 is the additive identity. For the sake of convenience, groups, rings, and fields will be denoted simply by G, R, F ,

respectively, whenever the binary operators are understood. The binary operator $+$ is called addition, and $*$ is called multiplication (or composition, or convolution). The product $a * b$ will usually be denoted simply by ab .

Suppose G and G' are two additive groupoids, and let $\alpha : G \rightarrow G'$. (If $G = G'$, then α is called an operator on G .) The function α is said to be additive (groupoid homomorphism) iff

$$(1) \quad \alpha(x + y) = \alpha x + \alpha y \quad (x, y \in G).$$

The concept of an additive function is very important in this paper.

If $\alpha : R \rightarrow R'$, where R and R' are rings, then α is said to be multiplicative iff

$$(2) \quad \alpha(xy) = (\alpha x)(\alpha y) \quad (x, y \in G),$$

and α is a ring homomorphism iff it is both additive and multiplicative.

If R, R' have multiplicative identities $1, 1'$ respectively and $\alpha 1 = 1'$,

then $\alpha(1x) = \alpha x = 1'(\alpha x) = (\alpha 1)(\alpha x)$; thus if α is only additive, it is

still true that (2) holds for some pairs x, y . If R and R' are commu-

tative rings, then clearly $\alpha(xy) = (\alpha x)(\alpha y)$ implies $\alpha(yx) = (\alpha y)(\alpha x)$,

and the relation $D_\alpha = [(x, y) \in G \times G : \alpha(xy) = (\alpha x)(\alpha y)]$ is symmetric.

The elements $x, y \in G$ will be called α -independent iff $\alpha(xy) = (\alpha x)(\alpha y)$.

For commutative rings, this notion can obviously be generalized to a finite number of terms, and then corresponds exactly to the concept of independence in probability.

Consider now an arbitrary nonempty set X , and let \mathcal{Q} be a non-empty class of subsets of X such that, if $A, B \in \mathcal{Q}$, then

$$(1) \quad A - B \in \mathcal{Q}$$

$$(2) \quad A \cup B \in \mathcal{Q}.$$

It follows that $\emptyset \in \mathcal{Q}$ and $AB = A \cup B - (AB^c + A^cB) \in \mathcal{Q}$. (Note: $A \cup B$ will be denoted by $A + B$ iff $AB = \emptyset$.) If the binary operation \pm is defined by

$$A \pm B = AB^c + A^cB = (A - B) + (B - A),$$

then it follows that $A \pm B = B \pm A$, $(A \pm B) \pm C = A \pm (B \pm C)$, and $(A \pm B)C = AC \pm BC$. Therefore (\mathcal{Q}, \pm, \cap) is a ring, where \emptyset is the zero element and the inverse of a set A is A itself. Conversely, if \mathcal{Q} is closed under \pm , \cap , and $-$, then it is closed under \cup , since $A \cup B = (A \pm B) + AB$. For this reason, \mathcal{Q} is called a (Boolean) ring of sets iff axioms (1) and (2) above are satisfied. A σ -ring is a nonempty class \mathcal{S} which is closed under differences and countable unions (hence also countable intersections).

If ζ is any nonempty class of sets, G is an additive Abelian group, and $\varphi : \zeta \rightarrow G$ is a set function, then φ is additive iff $\varphi(A + B) = \varphi A + \varphi B$ whenever $A, B \in \zeta$, $AB = \emptyset$, and $A + B \in \zeta$; φ is finitely additive iff $\varphi \sum_1^n A_k = \sum_1^n \varphi A_k$ whenever $A_k \in \zeta$, $A_j A_k = \emptyset$ ($j \neq k$), and $\sum_1^n A_k \in \zeta$. If \mathcal{Q} is a ring, and $\varphi : \mathcal{Q} \rightarrow G$ is additive, it follows that $\varphi(\emptyset) = \theta$, φ is finitely additive, $\varphi(A - B) = \varphi A - \varphi B$, and $\varphi(A \cup B) = \varphi A + \varphi B + \varphi AB$ whenever $A, B \in \mathcal{Q}$.

A linear space (vector space) is a pair (G, F) consisting of an additive Abelian group G , with zero element θ , and a field F of operators on G , with zero element 0 and unity element 1 , such that the following axioms are satisfied for $\alpha, \beta \in F$ and $x, y \in G$:

- (1) $\alpha(x + y) = \alpha x + \alpha y$ (the operators are additive)
- (2) $(\alpha + \beta)x = \alpha x + \beta x$ (operator addition law)

$$(3) (\alpha\beta)x = \alpha(\beta x) \quad (\text{operator composition law})$$

$$(4) 1x = x \quad (\text{the unity element } 1 \in F \text{ is the identity operator}).$$

It follows from (1) and (2) that $\alpha\theta = 0x = \theta$. The elements of G are called "vectors," and the elements of F are called "scalars." A linear space (X, F) will usually be denoted simply by X . In this paper, F will always be either the real line R^0 or the complex plane C^0 , and X will be called a real linear space or complex linear space accordingly. Note that if (X, C^0) is a complex linear space, then (X, R^0) is a real linear space, but not conversely; e.g. (C^0, C^0) is a complex linear space, (C^0, R^0) is a real linear space, and (R^0, R^0) is a real linear space, but (R^0, C^0) is not a linear space. From now on, the scalar field F will be denoted by Z , and may be either R^0 or C^0 unless otherwise specified.

If X and Y are two linear spaces with the same scalar field Z , and $T : X \rightarrow Y$ is a function, then T is a linear function iff

$$(1) T(x + y) = Tx + Ty \quad (T \text{ is } \underline{\text{additive}})$$

$$(2) T(\alpha x) = \alpha(Tx) \quad (T \text{ is } \underline{\text{homogeneous}}).$$

In case $X = Y$, T is called an operator, and in case $Y = Z$, T is called a functional. If $f : X \rightarrow Z$ and $f(x)$ is real for all x , then f is a sublinear functional iff

$$(1) f(x + y) \leq f(x) + f(y) \quad (f \text{ is } \underline{\text{subadditive}})$$

$$(2) f(\alpha x) = \alpha(fx) \text{ for } \alpha \geq 0 \quad (f \text{ is } \underline{\text{semihomogeneous}}).$$

If $Z = R^0$, then clearly every linear functional is sublinear.

A topology is a nonempty class \mathcal{U} such that

$$(1) U, V \in \mathcal{U} \triangleright UV \in \mathcal{U}$$

$$(2) V \subset \mathcal{U} \triangleright \bigcup_{V \in \mathcal{V}} V = \bigcup_{V \in \mathcal{V}} V \in \mathcal{U}$$

By taking the empty class for \mathcal{V} in (2), it follows that $\emptyset \in \mathcal{U}$, and by taking $\mathcal{V} = \mathcal{U}$ it follows that $\bigcup \mathcal{U} \in \mathcal{U}$; the pair (X, \mathcal{U}) is called a topological space. Suppose (Ω, \mathcal{I}) and (X, \mathcal{U}) are two topological spaces, and let $X : \Omega \rightarrow X$ be a function. Then X is said to be $(\mathcal{I}, \mathcal{U})$ -continuous iff

$$[X^{-1}U : U \in \mathcal{U}] \equiv X^{-1}\mathcal{U} \subset \mathcal{I}$$

A topological space (X, \mathcal{U}) will usually be denoted simply by X ; the sets U in the topology \mathcal{U} are called open sets, and

$$\mathcal{U}^c = [U^c : U \in \mathcal{U}]$$

is the class of all closed sets. It follows from DeMorgan's laws,

$$(\bigcap A_i)^c = \bigcup A_i^c, \quad (\bigcup A_i)^c = \bigcap A_i^c$$

that \mathcal{U}^c is closed under finite unions and arbitrary intersections.

Since $X^{-1}U^c = (X^{-1}U)^c$, it follows that a function $X : \Omega \rightarrow X$ is $(\mathcal{I}, \mathcal{U})$ -continuous iff $X^{-1}\mathcal{U}^c \subset \mathcal{I}^c$.

A much more detailed discussion of the topological concepts introduced below can be found in KELLEY.

The interior of a set A is the set

$$A^{\circ} = [B : A \supset B \in \mathcal{U}]$$

and is clearly the largest open set contained in A ; thus A is open iff $A = A^{\circ}$. It follows from the definition that $A^{\circ\circ} = A^{\circ}$, $(AB)^{\circ} = A^{\circ}B^{\circ}$, and $A \subset B \supset A^{\circ} \subset B^{\circ}$.

The closure of a set A is the set

$$A^- = \bigcap [C : A \subset C \in \mathcal{U}^c]$$

and is clearly the smallest closed set containing A : thus A is closed iff $A = A^-$. It follows that

$$A^{--} = A^-, \quad (A \cup B)^- = A^- \cup B^-, \text{ and } A \subset B \supset A^- \subset B^-.$$

A set V is a neighborhood (nhd.) of a point x iff $x \in V^0$, or equivalently iff there is an open set U such that $x \in U \subset V$. Note that a nhd. need not be open. The class \mathcal{N}_x of all neighborhoods of a point x is called the neighborhood system of x . It follows that a set S is open iff it is a nhd. of each of its points:

$$\mathcal{U} = [S : S \in \mathcal{N}_x \text{ for all } x \in S].$$

The following dual relations are consequences of the definitions of interior, closure, and neighborhood:

$$A^0 = [x : A^c V = \emptyset \text{ } (V \subset A) \text{ for some } V \in \mathcal{N}_x]$$

$$A^- = [x : AV \neq \emptyset \text{ for every } V \in \mathcal{N}_x]$$

$$A^{-c} = A^{c0}, \quad A^{oc} = A^{c-},$$

For each x , $\mathcal{N}'_x = [V' = V - x : V \in \mathcal{N}_x]$ is called the deleted nhd. system of x . Let

$$A^L = [x : AV' \neq \emptyset \text{ for every } V' \in \mathcal{N}'_x].$$

The points in A^L are called limit points of A , and it follows that A is closed iff $A^L \subset A$. Furthermore, $A^- = A \cup A^L$ for every set A .

A class $\mathcal{M}_x \subset \mathcal{N}_x$ is a local base at x , or a base for \mathcal{N}_x , iff

for every $N \in \mathcal{N}_x$ there is an $M \in \mathcal{M}_x$ such that $x \in M \subset N$.

A filter is a nonempty class \mathcal{F} of nonempty sets such that

$$(1) \quad A, B \in \mathcal{F} \quad A \cap B \in \mathcal{F}$$

$$(2) \quad A \supset B \in \mathcal{F} \quad A \in \mathcal{F}$$

A direction is a nonempty class \mathcal{D} of nonempty sets such that the intersection of two \mathcal{D} sets contains a \mathcal{D} set. It follows that every nhd. system \mathcal{N}_x is a filter, and every filter or local base is a direction.

A class \mathcal{J} is a base for a topology \mathcal{U} iff

$$\mathcal{U} = [\cup \mathcal{B} : \mathcal{B} \subset \mathcal{J}]$$

If \mathcal{U} is a given topology, it follows that \mathcal{J} is a base for \mathcal{U} iff

$\mathcal{J} \subset \mathcal{U}$ and $\mathcal{J}_x = [S : x \in S \in \mathcal{J}]$ is a base for \mathcal{N}_x for all x . On the

other hand, if \mathcal{J} is a given class, then the class $\mathcal{U} = [\cup \mathcal{B} : \mathcal{B} \subset \mathcal{J}]$ is a topology on X iff \mathcal{J}_x is a direction for all $x \in X$.

A topological space X is first countable (C_1) iff, for each $x \in X$, there is a countable base $\mathcal{M}_x = [M_n]$ for \mathcal{N}_x . If X is C_1 , it follows that for each x there is a countable local base $[M_n^x]$ such that $M_n^x \downarrow$; simply let $M_n^x = \bigcap_{k=1}^n M_k^x$.

A top. space X is second countable (C_2) iff there is a countable base $\mathcal{J} = [T_n]$ for the topology \mathcal{U} . Clearly every C_2 space is C_1 .

A set $D \subset X$ is dense in X iff $D^- = X$, and X is separable iff there is a countable dense set $D = [x_n] \subset X$. Every C_2 space is separable. Simply choose a point $x_n \in T_n$ for each n , where $[T_n]$ is a countable base for \mathcal{U} . Then for any x and any nhd U of x , $x_n \in T_n \subset U$ for some n ; thus $x \in [x_n]^-$.

Two sets A, B are said to be (weakly) separated iff there exist

open sets $U, V \in \mathcal{U}$ such that

$$A \subset UV^c \text{ and } B \subset U^cV.$$

The sum $A + B$ of two nonempty separated sets is called a separation, and a set S is connected iff it cannot be written as a separation. If $A + B \supset C$, where $A + B$ is a separation and C is connected, then either $A \supset C$ or $B \supset C$; for otherwise $AC + BC = C$ would be a separation.

Two sets A, B are strongly separated iff there exist open sets U, V such that $A \subset U$, $B \subset V$, and $UV = \emptyset$.

A top. space X is T_1 iff every two distinct points are weakly separated (equivalently $[x]$ is closed for all x); and X is Hausdorff (T_2) iff every two distinct points are strongly separated. Clearly every T_2 space is T_1 .

If X is a nonempty set, then a metric or distance function on X is a function $d : X \times X \rightarrow \mathbb{R}^+$ such that

$$(1) \quad d(x, y) \leq d(x, z) + d(y, z)$$

$$(2) \quad d(x, x) = 0$$

$$(3) \quad d(x, y) = 0 \supset x = y.$$

It follows from (1) and (2) that $d(x, y) = d(y, x) \geq 0$. If d satisfies only (1) and (2), then d is called a pseudo-metric (p-metric), and (X, d) is called a p-metric space; X is a metric space iff d is a metric. For each $x \in X$ and $r > 0$, the set

$$S(x, r) = [y : d(x, y) < r]$$

is called the open r-sphere centered at x . Let

$$\mathcal{J} = [S(x, r) : x \in X, r > 0].$$

It follows from the triangle inequality (1) that \mathcal{J} is a base for a topology \mathcal{U} on X , since $\mathcal{J}_x = [S \in \mathcal{J} : x \in S]$ is a direction for each $x \in X$. The topology \mathcal{U} is called the p-metric topology generated by d . A topological space (X, \mathcal{U}) is metrizable (p-metrizable) iff there exists a metric (p-metric) d such that \mathcal{U} is the metric (p-metric) topology generated by d . This metric (p-metric), if it exists, is not in general unique; in fact

$$d^0(x, y) = \frac{d(x, y)}{1 + d(x, y)} < 1, \quad d''(x, y) = \min[1, d(x, y)] \leq 1$$

are metrics (p-metrics) which both generate the same topology as d .

It follows easily that a p-metric space is metric iff the p-metric topology is T_1 . Every metric space is Hausdorff. Every p-metric space is regular: for each x , the class $\mathcal{N}_x \mathcal{U}^c$ of all closed nhds. of x is a base for \mathcal{N}_x .

The trivial topology on X is the topology $\mathcal{U}_0 = [\emptyset, X]$, and (X, \mathcal{U}_0) is a p-metric space with p-metric $d_0(x, y) = 0$. The discrete topology on X is the topology \mathcal{U}_1 such that $[x]$ is open for each $x \in X$ (i.e. $\mathcal{U}_1 = \mathcal{S}(X) = [\text{all subsets of } X]$), and is a metric space with metric $d_1(x, y) = 1$ ($x \neq y$), $d_1(x, x) = 0$.

Every p-metric space is C_1 , since for each x , $\mathcal{M}_x = [S(x, r) : r > 0, \text{ rational}]$ is a countable local base at x . A p-metric space X is C_2 if it is separable; for if $D = [x_n]$ is a countable dense set, then $\mathcal{J} = [S(x_n, q) : n = 1, 2, \dots, q > 0, \text{ rational}]$ is a countable base for the metric topology \mathcal{U} . To see this, take any point $x \in X$, and let U be any open set containing x and $S(x, r)$ an open sphere such that $S(x, r) \subset U$. Then $x_n \in S(x, r/2)$ for some n since $[x_n]$

is dense, and $x \in S(x_n, q) \subset S(x, r)$ for $d(x_n, x) < q < r/2$. Metric spaces will be called M_1 spaces, and separable metric spaces will be called M_2 spaces. Thus $M_1 \subset C_1$ and $M_2 \subset C_2$.

A class ζ is an open cover of a set S iff $\zeta \subset \mathcal{U}$ and $S \subset \bigcup \zeta$. A top. space X will be called L_1 iff every open cover of X has a countable subcover, and X will be called L_2 iff every open cover of an arbitrary set $S \subset X$ has a countable subcover. This terminology is used because of the following famous theorem.

Lindelöf Theorem. Every C_2 space is L_2 . (Thus $M_2 \subset C_2 \subset L_2$)

Proof. Let $\zeta \subset \mathcal{U}$ be an open cover of a set S , and let $[T_n]$ be a countable base for \mathcal{U} . For each $U \in \zeta$, write $U = \bigcup_k T_{n(k,u)}$. Then

$\mathcal{S} = \bigcup_{U \in \zeta} \bigcup_k [T_{n(k,u)}] = [T_{u_j}]$ is countable, where $T_{u_j} \subset U_j$ for $j = 1, 2, \dots$; hence $[U_j] \subset \zeta$ is a countable subcover, since $S \subset \bigcup T_{u_j} \subset \bigcup U_j$. □

A sequence $[x_n]$ in a top. space X is said to converge to a point x ($x_n \rightarrow x$) iff, for each $V \in \mathcal{N}_x$, there is a positive integer $p = p(x, V)$ such that $n \geq p \Rightarrow x_n \in V$. A sequence $[x_n]$ is said to be eventually in a set S iff there is a positive integer $p = p_S$ such that $n \geq p \Rightarrow x_n \in S$. Thus $x_n \rightarrow x$ iff x_n is eventually in every nhd. of x .

If X is C_1 , then the topology \mathcal{U} can be completely described in terms of sequential convergence:

$$A^0 = [x : x_n \rightarrow x \Rightarrow x_n \text{ is eventually in } A]$$

$$A^- = [x : x_n \rightarrow x \text{ for some sequence } x_n \in A].$$

Let (X, \mathcal{U}) and (Y, \mathcal{V}) be two top. spaces and $f : X \rightarrow Y$. If X is C_1 , then f is $(\mathcal{U}, \mathcal{V})$ -continuous iff

$$x_n \rightarrow x \triangleright f x_n \rightarrow f x.$$

To show this, suppose first that f is continuous, let $x_n \rightarrow x$, and choose any nhd. V of $f x$. Since f is continuous, there is a nhd. U of x such that $fU \subset V$. But x_n is eventually in U , hence $f x_n$ is eventually in $fU \subset V$. Now suppose f is not continuous. Then for some $x \in X$, there is a nhd. V of $f x$ such that $U = f^{-1}V$ is not a nhd. of x . Let $[M_n] \downarrow$ be a countable, decreasing local base at x , and for each n choose $x_n \in M_n^c$. Then $x_n \rightarrow x$, but $f x_n \notin V^c$ for all n , hence $f x_n \not\rightarrow f x$.

If X is a p -metric space, then clearly

$$x_n \rightarrow x \text{ iff } d(x_n, x) \rightarrow 0.$$

A sequence x_n in a p -metric space X is a Cauchy sequence iff $d(x_m, x_n) \rightarrow 0$ ($m, n \rightarrow \infty$), and X is complete iff every Cauchy sequence in X converges to some point in X . Complete metric spaces and separable complete metric spaces will be called \bar{M}_1 spaces and \bar{M}_2 spaces, respectively.

Note that limits are not in general unique in a p -metric space; but $x_n \rightarrow x$, $x_n \rightarrow y$ implies $d(x, y) = 0$. Clearly limits are unique in a Hausdorff space.

If X is a linear space with zero vector θ , then a norm on X is a function $|\cdot| : X \rightarrow \mathbb{R}^+$ such that

$$(1) \quad |\alpha x| = |\alpha| |x|$$

$$(2) \quad |x + y| \leq |x| + |y|$$

$$(3) \quad |x| = 0 \triangleright x = \theta.$$

Here $|\alpha|$ is the ordinary absolute value of α , and should not be confused with the norm of a vector, since α, β will always denote scalars and x, y vectors. It follows from (1) that $|\theta| = 0$, and then (1) and (2) with $y = -x$ imply that $|x| \geq 0$.

A normed linear space (NLS) is a pair $(X, |\cdot|)$ consisting of a linear space X and a norm on X . The function $d(x, y) = |x - y|$ is easily seen to be a metric, and generates a topology on X called the norm topology. It is easy to see that the function $T : X \rightarrow X$ defined by $Tx = \alpha x + a$, where $\alpha \neq 0$, is a homeomorphism of X onto X ; i.e. T is a 1-1 mapping of X onto itself such that T and T^{-1} are both continuous.

If $T : X \rightarrow Y$ is an additive transformation, then T is continuous on X iff it is continuous at the origin θ . For if $x_n \rightarrow \theta \triangleright Tx_n \rightarrow 0$, then $x_n \rightarrow x \triangleright x_n - x \rightarrow \theta$, hence

$$|Tx_n - Tx| = |T(x_n - x)| \rightarrow 0.$$

Similarly, a subadditive functional $f : X \rightarrow Z$ such that $f(\theta) = 0$ is continuous iff it is continuous at θ , since

$$-f(x - x_n) \leq f(x_n) - f(x) \leq f(x_n - x).$$

A field on X is a nonempty class $\zeta \subset \mathcal{B}(X)$ such that, if $A, B \in \zeta$, then

$$(1) A^c \in \zeta$$

$$(2) AB \in \zeta.$$

A σ -field on X is a nonempty class $\mathcal{B} \subset \mathcal{B}(X)$ such that, if $[A_n] \subset \mathcal{B}$, then

$$(1)_{\sigma} A_n^c \in \zeta$$

$$(2)_{\sigma} \bigcap A_n \in \zeta.$$

It follows by DeMorgan's laws that intersections can be replaced by unions in (2) and $(2)_{\sigma}$. Clearly every σ -field is a σ -ring and a field; and every field contains $\emptyset = AA^c$ and $X = \emptyset^c$, and is a ring.

Note that if ζ is a field, then (ζ, \pm, \cap) is a ring with unity element X , but is not a field in the algebraic sense, except in the trivial case $\zeta_0 = [\emptyset, X]$; since in all other cases X is the only set with an inverse with respect to intersection. The term "field" in this paper will refer to a field of sets from now on, and algebraic fields will be called "scalar fields."

If \mathcal{B} is a σ -field on X , then the pair (X, \mathcal{B}) is called a measurable space, and sets $B \in \mathcal{B}$ are called measurable sets, or events. Suppose (Ω, \mathcal{A}) and (X, \mathcal{B}) are two measurable spaces, and let $X : \Omega \rightarrow X$ be a function. Then X is called an $(\mathcal{A}, \mathcal{B})$ -measurable function iff

$$[X^{-1}B : B \in \mathcal{B}] \equiv X^{-1}\mathcal{B} \subset \mathcal{A}.$$

It follows easily that the intersection of any nonempty family of σ -fields is a σ -field. If ζ is any class of subsets of X , then there is at least one σ -field on X containing ζ ; namely the class $\mathcal{B}(X)$ of all subsets of X . By definition,

$$\bar{\zeta} = \bigcap [\mathcal{B} : \zeta \subset \mathcal{B}, \mathcal{B} \text{ is a } \sigma\text{-field}]$$

is the smallest σ -field containing ζ . It follows that ζ is a σ -field iff $\zeta = \bar{\zeta}$, and that

$$\zeta \subset \bar{\zeta}, \quad \bar{\bar{\zeta}} = \bar{\zeta}, \quad \text{and} \quad \zeta \subset \mathcal{D} \supset \bar{\zeta} \subset \bar{\mathcal{D}}.$$

Thus the operation of taking smallest σ -fields is similar to the operation of taking closures of sets in a topological space, except that the analogue of $(A \cup B)^- = A^- \cup B^-$ is not true since $\bar{\zeta} \cup \bar{\mathcal{D}}$ is not in general a σ -field. However it is true that $\overline{\zeta \cup \mathcal{D}} = \overline{\bar{\zeta} \cup \bar{\mathcal{D}}}$.

Now suppose $X : \Omega \rightarrow X$, where (Ω, \mathcal{A}) and (X, \mathcal{B}) are measurable spaces, and let ζ be any class such that $\mathcal{B} = \bar{\zeta}$. Then $X^{-1}\zeta \subset \mathcal{A}$ implies that X is $(\mathcal{A}, \mathcal{B})$ -measurable. For if $\mathcal{B}_1 = [B \in \mathcal{B} : X^{-1}B \in \mathcal{A}]$, then \mathcal{B}_1 is a σ -field and $\zeta \subset \mathcal{B}_1$, hence $\mathcal{B} = \bar{\zeta} \subset \bar{\mathcal{B}}_1 = \mathcal{B}_1$, which implies that $X^{-1}\mathcal{B} \subset \mathcal{A}$. Therefore, X is $(\mathcal{A}, \bar{\zeta})$ -measurable iff $X^{-1}\zeta \subset \mathcal{A}$.

A topological measurable (T.M.) space is a triple $(X, \mathcal{U}, \mathcal{B})$ such that (X, \mathcal{U}) is a topological space and (X, \mathcal{B}) is a measurable space. Suppose $(\Omega, \mathcal{J}, \mathcal{A})$ and $(X, \mathcal{U}, \mathcal{B})$ are two topological measurable spaces, and let $X : \Omega \rightarrow X$ be a function. Then X is $(\mathcal{J}, \mathcal{U})$ -continuous iff $X^{-1}\mathcal{U} \subset \mathcal{J}$ and $(\mathcal{A}, \mathcal{B})$ -measurable iff $X^{-1}\mathcal{B} \subset \mathcal{A}$. Because of the properties of inverse functions, it is clear that $\mathcal{J}_0 = X^{-1}\mathcal{U}$ is a topology and $\mathcal{A}_0 = X^{-1}\mathcal{B}$ is a σ -field; they are respectively the smallest topology \mathcal{J}_0 on Ω such that X is $(\mathcal{J}_0, \mathcal{U})$ -continuous and the smallest σ -field \mathcal{A}_0 on Ω such that X is $(\mathcal{A}_0, \mathcal{B})$ -measurable. Furthermore, $\mathcal{U}_1 = [U : X^{-1}U \in \mathcal{J}]$ is a topology on X and $\mathcal{B}_1 = [B : X^{-1}B \in \mathcal{A}]$ is a σ -field on X ; they are respectively the largest topology \mathcal{U}_1 on X such that X is $(\mathcal{J}, \mathcal{U}_1)$ -continuous and the largest σ -field \mathcal{B}_1 on X such that X is $(\mathcal{A}, \mathcal{B}_1)$ -measurable.

Let $X : \Omega \rightarrow X$ and $f : X \rightarrow Y$ where $(Y, \mathcal{V}, \mathcal{Z})$ is a third

T.M. space. If X is $(\mathcal{I}, \mathcal{U})$ -continuous or $(\mathcal{A}, \mathcal{B})$ -measurable, and f is $(\mathcal{U}, \mathcal{V})$ -continuous or $(\mathcal{B}, \mathcal{Z})$ -measurable, then it follows easily that the composition $fX : \Omega \rightarrow Y$ is $(\mathcal{I}, \mathcal{V})$ -continuous or $(\mathcal{A}, \mathcal{Z})$ -measurable, respectively.

In many applications, it is desirable that all continuous functions be measurable. Suppose X is continuous, so that $X^{-1}\mathcal{U} \subset \mathcal{I}$. If $\mathcal{A} \supset \mathcal{I}$ and $\mathcal{B} \subset \bar{\mathcal{U}}$, then X is measurable, since $X^{-1}\mathcal{U} \subset \mathcal{I}$ implies that $\mathcal{U} \subset \mathcal{B}_1 = [B : X^{-1}B \in \mathcal{A}]$, which implies that $\mathcal{B} \subset \bar{\mathcal{U}} \subset \mathcal{B}_1$, hence $X^{-1}\mathcal{B} \subset \mathcal{A}$. Therefore, $\mathcal{A} \supset \mathcal{I}$ and $\mathcal{B} \subset \bar{\mathcal{U}}$ imply that $(\mathcal{I}, \mathcal{U})$ -continuous functions are $(\mathcal{A}, \mathcal{B})$ -measurable.

In order to guarantee that compositions fX of continuous functions $X : \Omega \rightarrow X$ and $f : X \rightarrow Y$ be measurable, it therefore suffices to assume that $\mathcal{A} \supset \mathcal{I}$, $\mathcal{B} = \bar{\mathcal{U}}$, and $\mathcal{Z} \subset \bar{\mathcal{V}}$.

Suppose $[X_t : t \in T]$ is a nonempty family of nonempty sets, where T is a nonempty index set. By the Axiom of Choice, there exists a function $x : T \rightarrow \bigcup_{t \in T} X_t$ such that

$$x_t \equiv x(t) \in X_t, \quad t \in T.$$

The space of all such "choice" functions is denoted by $X^T = \prod_{t \in T} X_t$. Thus in particular if $X_t \equiv X$ for all $t \in T$, then $X^T = [x : T \rightarrow X]$. Note that the product space X^T is by definition a space of functions x , which can also be thought of as "vectors" $x = (x_t : t \in T)$, where the " t^{th} component" of x is $x_t \equiv x(t)$.

If $X = (X_t : t \in T)$ is a family of functions $X_t : \Omega \rightarrow X_t$, then X can also be thought of as a single "vector valued" function $X : \Omega \rightarrow X^T$, whose " t^{th} component function" is the function X_t ; thus

$X(\omega) = (X_t(\omega) : t \in T) \equiv (X(t, \omega) : t \in T)$ for each $\omega \in \Omega$. Then $X_t = p_t \circ X$, where p_t is the projection of X^T onto its t^{th} component space X_t , defined by $p_t x = x_t, t \in T$.

If S is a nonempty set, then a relation in S means a subset $< \subset S \times S$, and $a < b$ means $(a, b) \in <$. A partially ordered set is a pair $(S, <)$ consisting of a nonempty set S and a relation $<$ in S such that

$$(1) \quad a < b, \quad b < c \triangleright a < c$$

$$(2) \quad a < b, \quad b < a \text{ iff } a = b$$

A partially ordered set $(S, <)$ is said to be linearly ordered iff

$$(3) \quad a < b \text{ or } b < a \text{ for all pairs } a, b \in S.$$

If $(S, <)$ is a partially ordered set, $A \subset S$, and $b \in S$, then b is an upper bound for A iff $a < b$ for all $a \in A$; and b is a least upper bound for A iff it is an upper bound and $b < b'$ for any other upper bound b' . It follows from (2) that least upper bounds are unique whenever they exist. An element $m \in S$ is maximal iff

$$m < n \in S \triangleright m = n$$

A subset $C \subset S$ is called a chain iff $(C, <)$ is linearly ordered. The following famous lemma, which is equivalent to the Axiom of Choice, will be assumed as an axiom.

Zorn's Lemma. Let $(X, <)$ be a partially ordered set. If every chain C in S has an upper bound, then S has a maximal element.

A lattice is a partially ordered set $(S, <)$ such that every pair $a, b \in S$ has a least upper bound (supremum) $a \vee b$ and a greatest lower bound (infimum) $a \wedge b$. Thus \vee and \wedge are binary operators on

S such that

$$(1) \quad a < c, \quad b < c \quad \text{iff} \quad a \vee b < c$$

$$(2) \quad a > c, \quad b > c \quad \text{iff} \quad a \wedge b > c$$

where $a > c$ means $c < a$.

It is also easy to see that the following relations hold:

$$(1)' \quad a < b \quad \text{iff} \quad a \vee b = b, \quad (a \vee b) \vee c = a \vee (b \vee c)$$

$$(2)' \quad a < b \quad \text{iff} \quad a \wedge b = a, \quad (a \wedge b) \wedge c = a \wedge (b \wedge c)$$

In fact (1) and (1)' are equivalent, and (2) and (2)' are equivalent. The binary operators \vee and \wedge are called lattice operators. Note that (S, \vee) and (S, \wedge) are Abelian semigroups. Some examples of lattices are $(\mathcal{R}, \subset, \cup, \cap)$, and (R, \leq, \max, \min) , where \mathcal{R} is either a ring of sets or a topology, and R is the real line. These lattices are in fact distributive:

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c).$$

A σ -lattice is a lattice $(S, <)$ such that every sequence $[a_n] \subset S$ which is bounded above has a least upper bound $\vee a_n$, and every sequence $[b_n] \subset S$ which is bounded below has a greatest lower bound $\wedge b_n$. Thus if \mathcal{J}_a is the class of all sequences in S which are bounded above, and \mathcal{J}_b is the class of all sequences in S which are bounded below, then $\vee : \mathcal{J}_a \rightarrow S$ and $\wedge : \mathcal{J}_b \rightarrow S$ are operators such that

$$(1) \quad a_n < a' \quad \text{for all } n \quad \text{iff} \quad \vee a_n < a'$$

$$(2) \quad b_n > b' \quad \text{for all } n \quad \text{iff} \quad \wedge b_n > b'.$$

Examples of σ -lattices are $(\mathcal{R}, \subset, \cup, \cap)$ and (R, \leq, \sup, \inf)

where \mathcal{R} is a σ -ring and R is either the real line or the extended real line $\bar{R} = [-\infty, \infty]$.

One of the basic concepts which is used in this paper is the concept of convergence. The concept of sequential convergence in a topological space has already been introduced. But another notion of convergence also seems to appear frequently; namely convergence in a σ -lattice.

Suppose $(S, <)$ is a σ -lattice, and let $\mathcal{B} = \mathcal{B}_a \mathcal{B}_b$ be the class of all bounded sequences in S . Define $\bigwedge \bigvee a_n = \bigwedge_n \bigvee_{k \geq n} a_k$ and $\bigvee \bigwedge a_n = \bigvee_n \bigwedge_{k \geq n} a_k$ for $[a_n] \in \mathcal{B}$. Then since $\bigwedge_{j \geq m} a_j < \bigvee_{k \geq n} a_k$ for all m, n , it follows that $\bigvee_m \bigwedge_{j \geq m} a_j < \bigvee_n \bigwedge_{k \geq n} a_k$ for all n , hence

$$\bigvee \bigwedge a_m = \bigvee_m \bigwedge_{j \geq m} a_j < \bigwedge_n \bigvee_{k \geq n} a_k = \bigvee \bigwedge a_n.$$

By definition, a sequence $[a_n] \in \mathcal{B}$ is said to converge to an element $a \in S$ ($a_n \rightarrow a$, $\lim a_n = a$) iff

$$\bigwedge \bigvee a_n = \bigvee \bigwedge a_n = a.$$

Note that, to show that a sequence $[a_n] \in \mathcal{B}$ converges, it suffices to show that $\bigwedge \bigvee a_n < \bigvee \bigwedge a_n$. In particular if $[a_n] \in \mathcal{B}$ and $a_n \uparrow$ ($a_n < a_{n+1}$ for all n), then $a_n = \bigwedge_{k \geq n} a_k$, thus $\bigvee_{j \geq m} a_j = \bigvee a_n = \bigvee (\bigwedge_{k \geq n} a_k)$ implies $\bigwedge \bigvee a_n = \bigvee a_n = \bigvee \bigwedge a_n$; similarly, if $a_n \downarrow$ ($a_n > a_{n+1}$), then $\lim a_n = \bigwedge a_n$. If $a_n = a$ for all n , then clearly $\lim a_n = a$.

A function $\phi : S \rightarrow X$, where S is a σ -lattice and X is either a topological space or a σ -lattice, is said to be continuous at a point $a \in S$ iff

$$a_n \rightarrow a \text{ in } S \triangleright \varphi(a_n) \rightarrow \varphi(a) \text{ in } X.$$

If $a_n \uparrow a \triangleright \varphi(a_n) \rightarrow \varphi(a)$, then φ is said to be continuous from below at a , and if $a_n \downarrow a \triangleright \varphi(a_n) \rightarrow \varphi(a)$, then φ is said to be continuous from above at a .

If $\varphi : S \rightarrow X$, where S and X are both σ -lattices, then φ is said to be monotone (increasing) iff $a < b$ in $S \triangleright \varphi(a) < \varphi(b)$ in X . It follows that a monotone function is continuous iff it is continuous from above and from below. To show this, let $a_n \rightarrow a$, and note that

$$a'_n = \bigvee_{k \geq n} a_k \downarrow a \text{ and } \bigwedge_{k \geq n} a_k \uparrow a. \text{ Hence,}$$

$$a_n < a'_n \triangleright \varphi(a_n) < \varphi(a'_n) \triangleright \bigwedge \varphi(a_n) < \bigwedge \varphi(a'_n) = \varphi(a)$$

by continuity from above. Similarly, $\bigvee \varphi(a_n) > \varphi(a)$, hence

$$\bigwedge \varphi(a_n) < \varphi(a) < \bigvee \varphi(a_n).$$

This concept of continuity is very important in later chapters.

CHAPTER III

MEASURES AND INTEGRALS

In this chapter, the basic parts of measure theory and integration which are used in probability theory will be developed. For a more complete discussion of this material, see HALMOS [1], LOEVE, and HILLE and PHILLIPS. Throughout the chapter, (Ω, \mathcal{A}) is an arbitrary measurable space; and X is a NLS with scalar field Z , norm topology \mathcal{U} , and Borel σ -field $\mathcal{B} = \bar{\mathcal{U}}$. In many cases $X = Z = \mathbb{R}$ or \mathbb{C} .

As noted in the preceding chapter, the concepts of additive function and continuous functions on a σ -lattice are extremely important.

Let $\phi : \zeta \rightarrow X$ be a set function on a class $\zeta \subset \mathcal{B}(\Omega)$ containing the empty set. Then ϕ is σ -additive iff

$A_n \in \zeta, \sum A_n \in \zeta (A_m A_n = \emptyset, m \neq n) \Rightarrow \phi \sum A_n = \sum \phi A_n$. If ϕ is either additive or σ -additive, it follows that $\phi(\emptyset) = \theta$. If ϕ is σ -additive, then it is called an X -valued measure on ζ . Only the case $X = Z$ will be considered in this chapter; in this case a scalar valued measure ϕ is called a real measure or a complex measure according as $Z = \mathbb{R}$ or $Z = \mathbb{C}$. A real measure ϕ is a measure iff $\phi \geq 0$. If $\phi \geq 0$ and ϕ is finitely additive, then ϕ is called a content. If $\Omega \in \zeta$, then a probability on ζ is a measure P on ζ with $P\Omega = 1$.

In the case of real measures, the value ∞ is sometimes allowed, but never the value $-\infty$.

Let $\phi : \zeta \rightarrow \bar{\mathbb{R}}$ be a set function on a class ζ such that $-\infty < \phi A \leq \infty$, and consider $A_n \in \zeta$. Then ϕ is said to be

- (1) continuous from below iff $A_n \uparrow A \in \zeta \triangleright \varphi A = \lim \varphi A_n$
- (2) continuous from above iff $A_n \downarrow A \in \zeta, \varphi A_1 < \infty \triangleright \varphi A = \lim \varphi A_n$
- (3) continuous iff φ is continuous from above and from below.

Note that if $\varphi A < \infty$ for all $A \in \zeta$, φ is monotone, and ζ is a σ -ring (hence also a σ -lattice), then this definition of continuity is equivalent to the definition of a continuous function on a σ -lattice given in the previous chapter.

Suppose now that ζ is a nonempty class which is closed under finite sums and proper differences; so that if $A_1, A_2 \in \zeta$, then $A_1 A_2 = \emptyset \triangleright A_1 + A_2 \in \zeta$ and $A_1 \supset A_2 \triangleright A_1 - A_2 \in \zeta$. In particular this will be true if ζ is a ring. The following theorem is very important, and is probably used in this paper more than any other single theorem.

Continuity Theorem. Let ζ be a class which is closed under finite sums and proper differences, and suppose $\varphi : \zeta \rightarrow (-\infty, \infty]$ is additive.

(a) Then σ -additivity is equivalent to continuity from below, and continuity from below implies continuity from above.

(b) If φ is finite, then φ is continuous (σ -additive) iff it is continuous at \emptyset .

Proof. (a) Suppose first that φ is continuous from below, and let $A = \sum A_n \in \zeta$, $B_n \downarrow B \in \zeta$, where $A_n, B_n \in \zeta$ and $\varphi B_1 < \infty$ (hence $\varphi B_n < \infty$ for all n).

Then $\sum_1^n A_k \uparrow A$ implies $\sum_1^n \varphi A_k \rightarrow \varphi A$; and $B_1 - B_n \uparrow B_1 - B$ implies $\varphi B_1 - \varphi B_n \rightarrow \varphi B_1 - \varphi B$, which implies $\varphi B_n \rightarrow \varphi B$. Therefore φ is σ -additive and continuous from above.

Now suppose φ is σ -additive, and let $A_n \uparrow A \in \mathcal{Z}$ and $A_0 = \emptyset$. Then $A = \Sigma(A_n - A_{n-1})$ implies $\varphi A = \Sigma \varphi(A_n - A_{n-1}) = \lim \Sigma_1^n (\varphi A_k - \varphi A_{k-1}) = \lim \varphi A_n$, hence φ is continuous from below.

(b) Suppose φ is finite and continuous at \emptyset , and let $A_n \uparrow A \in \mathcal{Z}$, $B_n \downarrow B \in \mathcal{Z}$. Then $A - A_n \downarrow \emptyset \triangleright \varphi A - \varphi A_n \rightarrow 0 \triangleright \varphi A_n \rightarrow \varphi A$, and $B_n - B \downarrow \emptyset \triangleright \varphi B_n - \varphi B \rightarrow 0 \triangleright \varphi B_n \rightarrow \varphi B$. ■

Note that part (b) is analogous to the fact that an additive transformation on a vector space is continuous iff it is continuous at the origin θ .

The following lemma depends strongly on the Continuity Theorem, and is in turn used to prove the Hahn-Jordan Theorem, which is used to prove the Radon-Nikodym Theorem.

Maximum-minimum lemma. Every real measure φ on a σ -field \mathcal{A} assumes a maximum and a minimum on \mathcal{A} .

Outline of Proof. It suffices to assume that $0 < \sup \varphi$ and $\varphi < \infty$; otherwise the proof is trivial. Choose $A_n \in \mathcal{A}$ so that $0 < \varphi A_n \rightarrow \sup \varphi$, and let $A = \bigcup A_n$. It follows by induction that for each n , A can be partitioned into 2^n disjoint sets, say $A = \sum_{m=1}^{2^n} A_{nm}$, where

$A_{nm} = \bigcap_{k=1}^n A_{nmk}$ and each A_{nmk} is either A_k or $A - A_k$. Note that

$A_n = \Sigma[A_{nm} : A_{nmn} = A_n, 1 \leq m \leq 2^n]$.

Let $B_n = \Sigma[A_{nm} : \varphi A_{nm} > 0, 1 \leq m \leq 2^n]$. It follows that

$\varphi A_n \leq \varphi B_n \leq \varphi \bigcup_{k=n}^{n'} B_k \uparrow \varphi \bigcup_{k=n}^{\infty} B_k$ by continuity from below, then

$\sup \varphi \leq \varphi \bigcap_n \bigcup_{k \geq n} B_k$ by continuity from above, since φ is finite.

Thus the maximum is attained, and similarly there is a set $D \in \mathcal{A}$ such that $-\varphi D = \sup (-\varphi)$, hence $\varphi D = \inf \varphi$. ■

Hahn-Jordan Decomposition Theorem. If φ is a real measure on a σ -field \mathcal{A} , then there is a set $D \in \mathcal{A}$ such that, for all $A \in \mathcal{A}$,

$$\varphi(AD^c) = \sup_{A \supset B \in \mathcal{A}} \varphi B \equiv \varphi^+ A, \quad \varphi(AD) = \inf_{A \supset B \in \mathcal{A}} \varphi B \equiv -\varphi^- A.$$

Outline of Proof. By the max-min lemma there is a set $D \in \mathcal{A}$ such that $\varphi D = \inf \varphi > -\infty$. Thus

$$\varphi(D - AD) = \varphi D - \varphi AD \geq \varphi D, \quad \varphi(D + AD^c) = \varphi D + \varphi AD^c \geq \varphi D$$

which implies $-\varphi AD \geq 0$ and $\varphi AD^c \geq 0$. Thus if $A \supset B$, then

$$\varphi B \leq \varphi BD^c \leq \varphi BD^c + \varphi AB^c D^c = \varphi AD^c,$$

which means $\varphi^+ A \leq \varphi AD^c$, hence $\varphi^+ A = \varphi AD^c$. Similarly $\varphi AD = -\varphi^- A$, hence $\varphi = \varphi^+ - \varphi^-$. ■

Since φ^+ and φ^- are measures, this means that every real measure is the difference of two measures. This result is analogous (and closely related) to the result that every function of bounded variation is the difference of two increasing functions.

The Borel σ -field \mathcal{B} on the real line R is the minimal σ -field containing all open sets: $\mathcal{B} = \bar{\mathcal{U}}$, where \mathcal{U} is the usual topology on R . Since every open set in R is a countable union of open intervals, it follows easily that \mathcal{B} is the minimal σ -field containing the class \mathcal{I} of

all open intervals. If $\zeta = [[a, b) : a, b \in \mathbb{R}]$, $\zeta_1 = [(r, \infty) : r \in \mathbb{R}]$, and $\zeta_2 = [(-\infty, r) : r \in \mathbb{R}]$, then it follows easily that

$$\mathcal{B} = \bar{\mathcal{U}} = \bar{\mathcal{J}} = \bar{\zeta} = \bar{\zeta}_1 = \bar{\zeta}_2.$$

Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and recall that a function $X : \Omega \rightarrow \mathbb{R}$ is $(\mathcal{A}, \mathcal{B})$ measurable iff $X^{-1}\mathcal{B} \subset \mathcal{A}$. Let \mathcal{M} denote the class of all $(\mathcal{A}, \mathcal{B})$ measurable functions. Note that the class \mathcal{R}^Ω of all functions $X : \Omega \rightarrow \mathbb{R}$ is a σ -lattice, with $X \leq Y$ iff $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$. It will now be shown that \mathcal{M} is a sub- σ -lattice of \mathcal{R}^Ω .

Suppose $[X_n] \subset \mathcal{M}$ and $-\infty < \bigwedge X_n \leq \bigvee X_n < \infty$, and let $X = \bigvee X_n = \sup X_n$. Then for each $r \in \mathbb{R}$,

$$X^{-1}(r, \infty) = \bigcup_k [\omega : \sup_n X_n > r + \frac{1}{k}] = \bigcup_m X_m^{-1}(r, \infty) \in \mathcal{A}$$

since $X_m \in \mathcal{M}$ for each m , and this means $X \in \mathcal{M}$ since $\mathcal{B} = \bar{\zeta}_1$. Similarly it follows that $\bigwedge X_n \in \mathcal{M}$, hence $\limsup X_n = \bigwedge_n \bigvee_{k \geq n} X_k \in \mathcal{M}$ and $\liminf X_n = \bigvee_n \bigwedge_{k \geq n} X_k \in \mathcal{M}$, which means that $\lim X_n \in \mathcal{M}$ whenever it exists and is finite.

Functions of the form $X = \sum_1^m x_j 1_{A_j}$, where $x_j \in \mathbb{R}$ and $A_j \in \mathcal{A}$ for $j = 1, \dots, m$ are called (real) simple functions.

Suppose $X \in \mathcal{M}$, and for each $n = 1, 2, \dots$, let

$$X_n = \sum_{k=-n2^n}^{n2^n} \frac{k-1}{2^n} 1 \left[\frac{k-1}{2^n} \leq X < \frac{k}{2^n} \right]$$

Then clearly each X_n is a simple function, since $X \in \mathcal{M}$ implies

$$A_{nk} = X^{-1} \left[\frac{k-1}{2^n}, \frac{k}{2^n} \right) \in \mathcal{A}. \text{ Furthermore, it is clear that}$$

$X_n(\omega) \rightarrow X(\omega)$ for all $\omega \in \Omega$. In fact, if $X \geq 0$, then all terms for $k \leq 0$ vanish and $0 \leq X_n \uparrow X$. That is, every nonnegative measurable function is the limit of a nondecreasing sequence of nonnegative simple functions.

Two functions $X, Y \in \mathcal{M}$ are said to be equal almost surely (a.s.) (or almost everywhere), with respect to μ , written $X \stackrel{\text{a.s.}}{=} Y$, iff $\mu[X \neq Y] = 0$. It follows easily that $\stackrel{\text{a.s.}}{=}$ is an equivalence relation, which therefore partitions \mathcal{M} into equivalence sets. It will often be convenient to identify functions which are a.s. equal.

If $X : \Omega \rightarrow \mathbb{R}$ is the limit of a sequence of simple functions, then $X \in \mathcal{M}$, since the limit of a sequence of measurable functions is measurable. Therefore, $X \in \mathcal{M}$ if and only if there is a sequence of simple functions which converges to X . This implies at once that, if $X, Y \in \mathcal{M}$, then αX , $X + Y$, and XY are also in \mathcal{M} . Thus \mathcal{M} is a linear space, and also a ring (algebra).

Let \mathcal{M}^+ denote the class of all nonnegative measurable functions, and \mathcal{M}_s^+ the class of all nonnegative simple functions, and suppose $X, Y \in \mathcal{M}_s^+$, say

$$X = \sum_1^m x_j I_{A_j}, \quad Y = \sum_1^n y_k I_{B_k},$$

where $x_j, y_k \geq 0$, $A_j, B_k \in \mathcal{A}$, and $\Omega = \sum_1^m A_j = \sum_1^n B_k$. Then,

$$X = \sum_1^m \sum_1^n x_j I_{A_j B_k}, \quad \text{and} \quad Y = \sum_1^n \sum_1^m y_k I_{A_j B_k}.$$

Note that if $X = Y$ and $A_j B_k \neq \emptyset$, then $x_j = y_k$, hence

$$\sum_1^m x_j \mu A_j = \sum_1^m \sum_1^n x_j \mu A_j B_k = \sum_1^n \sum_1^m y_k \mu A_j B_k = \sum_1^n y_k \mu B_k.$$

Definition. $\int X d\mu = \int X = \sum_1^m x_j \mu A_j$ if $X = \sum_1^m x_j 1_{A_j} \in \mathcal{M}_s^+$. It follows easily that, if $X, Y \in \mathcal{M}_s^+$, then $\int \alpha X = \alpha \int X$, $\int (X+Y) = \int X + \int Y$, and $X \leq Y \triangleright \int X \leq \int Y$. For the sake of convenience, $\int X d\mu$ will usually be denoted by $\int X$ whenever the measure μ is understood.

The proof of the following lemma is again based on the Continuity Theorem, and justifies the crucial step in the definition of the integral.

Monotone Convergence Lemma. If $X_n, Y \in \mathcal{M}_s^+$ and $X_n \uparrow X \geq Y$, then

$$\lim \int X_n \geq \int Y.$$

Proof. Consider first the case $Y = IB$, where $B \in \mathcal{A}$. Let $\epsilon > 0$ and $E_n = [\omega \in B : X_n(\omega) > 1 - \epsilon]$. Then $E_n \uparrow B$, hence $\mu E_n \uparrow \mu B$ by the Continuity Theorem. Thus $\int X_n \geq \int_{E_n} X_n \geq (1 - \epsilon) \mu E_n$ implies $\lim \int X_n \geq (1 - \epsilon) \mu B$, which implies $\lim \int X_n \geq \mu B = \int Y$ since ϵ is arbitrary.

Now if $Y = \sum_1^m y_k IB_k$, where $\Omega = \sum_1^m B_k$, then $\lim X_n \geq Y$ implies $\lim_n X_n IB_k \geq y_k IB_k$ for $k = 1, \dots, m$, hence $\lim_n \int X_n IB_k \geq y_k \mu B_k$. But $\int X_n = \sum_1^m \int X_n IB_k$, hence

$$\lim_n \int X_n = \sum_1^m \lim_n \int X_n IB_k \geq \sum_1^m y_k \mu B_k = \int Y. \quad \blacksquare$$

This implies that if $X_n, Y_n \in \mathcal{M}_s^+$ and $X_n, Y_n \uparrow X$, then

$\lim \int X_n = \lim \int Y_n$. For $X_m \uparrow X \geq Y_n$ implies $\lim \int X_m \geq \int Y_n$ for all n , hence $\lim \int X_m \geq \lim \int Y_n$, and similarly the opposite inequality holds.

Definition. If $X \in \mathcal{M}^+$, then $\int X = \lim \int X_n$, where $X_n \in \mathcal{M}_s^+$ and $X_n \uparrow X$.

If $X \in \mathcal{M}$, write $X = X^+ - X^-$, where $X^+ = \sup(X, 0)$ and $X^- = -\inf(X, 0)$. If either $\int X^+ < \infty$ or $\int X^- < \infty$, define $\int X = \int X^+ - \int X^-$. Note that $|\int X| \leq \int X^+ + \int X^- = \int (X^+ + X^-) = \int |X|$, and that $|\int X| < \infty$ iff $\int |X| < \infty$. The class of all $X \in \mathcal{M}$ such that $\int |X| < \infty$ will be denoted by \mathcal{L} , and functions $X \in \mathcal{L}$ will be called

Lebesgue summable.

The following four theorems are among the most celebrated and powerful theorems in real analysis.

Monotone Convergence Theorem. If $X_n \in \mathcal{M}^+$ and $X_n \uparrow X$, then $\int X_n \uparrow \int X$.

Proof. For each n , there is a sequence $[X_{nk}]$ of simple functions such that $0 \leq X_{nk} \uparrow X_n (k \rightarrow \infty)$. Let $Y_n = \sup [X_{mk} : 1 \leq m \leq n, 1 \leq k \leq n]$. Then $Y_n \in \mathcal{M}_s^+$ and $X_n \geq Y_n \uparrow X$, therefore

$$\lim \int X_n \geq \lim \int Y_n = \int X$$

by the definition of $\int X$. But clearly $\lim \int X_n \leq \int X$, hence

$$\lim \int X_n = \int X.$$

The following famous lemma is a corollary of the Monotone

Convergence Theorem (MCT). The notation $\underline{\lim} = \liminf$ and $\overline{\lim} = \limsup$ will be used.

Fatou's Lemma. If $X_n, Y \in \mathcal{M}^+$ and $\int Y < \infty$, then

$$(a) \quad \underline{\lim} \int X_n \geq \int \underline{\lim} X_n$$

$$(b) \quad \overline{\lim} \int X_n \leq \int \overline{\lim} X_n \quad \text{if } X_n \leq Y.$$

Proof. (a) Let $\underline{X}_n = \inf_{k \geq n} X_k$. Then $X_n \geq \underline{X}_n \uparrow \underline{X} = \underline{\lim} X_n$, hence

$\int X_n \geq \int \underline{X}_n \uparrow \int \underline{X}$ by the MCT, which implies that $\underline{\lim} \int X_n \geq \int \underline{X}$.

(b) Let $\overline{X}_n = \sup_{k \geq n} X_k$. Then $X_n \leq \overline{X}_n \downarrow \overline{X} = \overline{\lim} X_n$, hence

$0 \leq \overline{X}_1 - \overline{X}_n \uparrow \overline{X}_1 - \overline{X}$. This implies that $\int \overline{X}_1 - \int \overline{X}_n \uparrow \int \overline{X}_1 - \int \overline{X}$,

hence $\int X_n \leq \int \overline{X}_n \downarrow \int \overline{X}$, using the MCT and finiteness of the integrals.

Thus $\overline{\lim} \int X_n \leq \int \overline{X}$. ■

Dominated Convergence Theorem (Lebesgue DCT). If $X_n \in \mathcal{M}$,

$|X_n| \leq Y \in \mathcal{L}$, and $X_n \rightarrow X$, then $X \in \mathcal{L}$ and $\int X_n \rightarrow \int X$.

Proof. By hypothesis, $X_n^+ \rightarrow X^+$, $X_n^- \rightarrow X^-$, and $X_n^+, X_n^- \leq Y$. Therefore, by Fatou's lemma,

$$\overline{\lim} \int X_n^+ \leq \int X^+ \leq \underline{\lim} \int X_n^+ < \infty$$

$$\overline{\lim} \int X_n^- \leq \int X^- \leq \overline{\lim} \int X_n^- < \infty.$$

Thus $\int X = \int X^+ - \int X^- \leq \underline{\lim} \int X_n^+ - \overline{\lim} \int X_n^- = \underline{\lim} \int X_n^+ + \underline{\lim} (-\int X_n^-) \leq$

$\underline{\lim} (\int X_n^+ - \int X_n^-) = \underline{\lim} \int X_n$. Similarly $\overline{\lim} \int X_n \leq \int X$, therefore

$$\lim \int X_n = \int X.$$

The MCT and DCT remain valid if $X_n \rightarrow X$ is replaced by $X_n \xrightarrow{a.s.} X$; that is $\mu[X_n \nrightarrow X] = 0$.

If μ is a measure and φ is a scalar measure, then φ is said to be μ -continuous iff

$$\mu A = 0 \quad \triangleright \quad \varphi A = 0.$$

Radon-Nikodym Theorem. Let μ be a σ -finite measure on the σ -field \mathcal{A} , and suppose $\varphi : \mathcal{A} \rightarrow \mathbb{C}$ is a σ -finite scalar measure. Then φ is μ -continuous iff there exists a function $Y \in \mathcal{L}$, determined up to a μ equivalence, such that

$$\varphi A = \int_A Y d\mu, \quad A \in \mathcal{A}.$$

By definition, $\frac{d\varphi}{d\mu} = Y$.

The sufficiency follows from the MCT. For the proof of the necessity, see LOEVE.

Suppose now that $X = X^r + iX^i : \Omega \rightarrow \mathbb{C}$, where $X^r, X^i \in \mathcal{M}$. The class of all such functions will be denoted by $\mathcal{M}_{\mathbb{C}}$. If $X^r, X^i \in \mathcal{L}$, then by definition $\int X = \int X^r + i \int X^i$, and the class of all these functions is denoted by $\mathcal{L}_{\mathbb{C}}$. It follows easily that $\mathcal{M}_{\mathbb{C}}$ and $\mathcal{L}_{\mathbb{C}}$ are linear spaces, and $\mathcal{L}_{\mathbb{C}}$ is a NLS with norm defined by

$$\|X\| = \int |X|.$$

Writing $\int X = \left| \int X \right| e^{i\alpha}$, it follows that $\left| \int X \right| = \int X e^{-i\alpha} \leq \int |X|$.

The Dominated Convergence Theorem extends immediately to the case of complex valued functions, for if $|X_n| \leq Y \in \mathcal{L}$, then $|X_n^r|, |X_n^i| \leq Y$; so that $X_n \rightarrow X$ implies $\lim \int X_n^r = \int X^r$ and $\lim \int X_n^i = \int X^i$.

If $X \in \mathcal{M}^+$ and $\int X = 0$, then $X \stackrel{a.s.}{=} 0$. That is, $\mu[X > 0] = 0$, or equivalently $\mu[X > \frac{1}{n}] = 0$ for every n by the Continuity Theorem, since

$$[X > \frac{1}{n}] \uparrow [X > 0].$$

For if $\mu[X > \frac{1}{n}] = a_n > 0$ for some n , then

$$\int_{[X > \frac{1}{n}]} X \geq \int_{[X > \frac{1}{n}]} \frac{1}{n} \mu[X > \frac{1}{n}] = \frac{1}{n} a_n,$$

which is a contradiction.

Thus if $X \in \mathcal{L}$ and $\int_A X = 0$ for every $A \in \mathcal{A}$, then

$$\int_{[X \geq 0]} X = 0 \text{ implies } X^+ \stackrel{a.s.}{=} 0, \text{ and}$$

$$\int_{[X \leq 0]} X = 0 \text{ implies } X^- \stackrel{a.s.}{=} 0,$$

hence $X = X^+ - X^- \stackrel{a.s.}{=} 0$. These simple facts are often useful.

Schwarz Inequality. If $|X|^2, |Y|^2 \in \mathcal{L}$, then

$$\left(\int |XY| \right)^2 \leq \left(\int |X|^2 \right) \left(\int |Y|^2 \right).$$

Proof. $\int (|X| + \lambda |Y|)^2 = \int |X|^2 + 2\lambda \int |XY| + \lambda^2 \int |Y|^2 \geq 0$ for all $\lambda \in \mathbb{R}$. \square

Markov-Tchebychev Inequality. If $X \in \mathcal{M}$, $r \in \mathbb{R}$ and $\epsilon > 0$, then

$$\mu[|X| \geq \epsilon] \leq \epsilon^{-r} \int |X|^r$$

Proof. $\int |X|^r \geq \int |X|^r \mathbb{I}[|X| \geq \epsilon] \geq \epsilon^r \mu[|X| \geq \epsilon]$. ■

The expectation of a vector valued random variable is defined as its Bochner integral. Before defining the Bochner integral, however, the following convergence theorems are needed. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and $X_n, X : \Omega \rightarrow X$. Then

(a) $X_n \xrightarrow{a.s.} X$ iff there exists a null set N such that

$$X_n \rightarrow X \text{ on } N^c,$$

(b) $X_n \xrightarrow{a.u.} X$ (almost uniformly) iff for every $\epsilon > 0$ there is a set $A_\epsilon \in \mathcal{A}$ such that $\mu A_\epsilon < \epsilon$ and

$$X_n \xrightarrow{u} X \text{ (uniformly) on } A_\epsilon^c,$$

(c) $X_n \xrightarrow{\mu} X$ iff for every $\epsilon > 0$

$$\mu[\omega : |X_n(\omega) - X(\omega)| \geq \epsilon] \rightarrow 0.$$

Convergences Theorem. Let $X_n, X : \Omega \rightarrow X$ be functions such that

$$|X_m - X|, |X_m - X_n| \in \mathcal{M}$$

(a) μ finite $\triangleright X_n \xrightarrow{a.s.} X$ iff $\lim_n \mu \bigcup_{m \geq n} [|X_m - X| \geq \epsilon] = 0$ for every $\epsilon > 0$,

(b) $X_m - X_n \xrightarrow{a.s.} 0$ iff $\mu \bigcap_n \bigcup_{m \geq n} [|X_m - X_n| \geq \epsilon] = 0$ for every $\epsilon > 0$,

(c) if $X_m - X_n \xrightarrow{\mu} 0$, then there is a subsequence, say X'_n ,

such that $X'_m - X'_n \xrightarrow{a.s.} 0$.

Proof. By definition, $X_n \xrightarrow{a.s.} X$ iff $\mu[X_n \not\rightarrow X] = 0$. But

$$\begin{aligned} [X_n \rightarrow X] &= \bigcap_k [\text{for some } n, |X_m - X| < \frac{1}{k} \text{ for all } m \geq n] \\ &= \bigcap_k \bigcup_n \bigcap_{m \geq n} [|X_m - X| < \frac{1}{k}]. \end{aligned}$$

Therefore, by DeMorgan's laws,

$$[X_n \not\rightarrow X] = \bigcup_k \bigcap_n \bigcup_{m \geq n} [|X_m - X| \geq \frac{1}{k}].$$

By the Continuity Theorem it follows that

$$\mu[X_n \not\rightarrow X] = \lim_k \mu \bigcap_n \bigcup_{m \geq n} [|X_m - X| \geq \frac{1}{k}] = 0 \text{ iff}$$

$$\mu \bigcap_n \bigcup_{m \geq n} [|X_m - X| \geq \frac{1}{k}] = 0 \text{ for every } k.$$

But applying the Continuity Theorem again, the above is equal to

$$\lim_n \mu \bigcup_{m \geq n} [|X_m - X_n| \geq \frac{1}{k}], \text{ and (a) follows.}$$

The proof of (b) is similar to the first part of (a).

Now suppose $X_j - X_k \xrightarrow{\mu} 0$, so that $\mu[|X_j - X_k| \geq 2^{-n}] \rightarrow 0$ ($j, k, \rightarrow \infty$) for each n . Choose q_n so that $q_n \uparrow \infty$ and

$$j, k \geq q_n \triangleright \mu[|X_j - X_k| \geq 2^{-n}] < 2^{-n}$$

and let $Y_n = X_{q_n}$.

Given $\epsilon > 0$, consider n sufficiently large so that $2^{-n+1} < \epsilon$.

It follows that

$$\bigcap_{m \geq n} [|Y_{m+1} - Y_m| < 2^{-m}] \subset \bigcap_{m \geq n} [|Y_m - Y_n| < \epsilon]$$

since $|Y_m - Y_n| \leq \sum_{m \geq n} |Y_{m+1} - Y_m|$ for $m \geq n$.

$$\text{Thus } \mu \bigcap_n \bigcup_{m \geq n} [|Y_m - Y_n| \geq \epsilon] \leq \mu \bigcup_{m \geq n} [|Y_m - Y_n| \geq \epsilon] \leq$$

$$\begin{aligned} & \mu \bigcup_{m \geq n} [|Y_{m+1} - Y_m| \geq 2^{-m}] \leq \sum_{m \geq n} \mu [|Y_{m+1} - Y_m| \geq 2^{-m}] \\ & \leq \sum_{m \geq n} 2^{-m} = 2^{-n+1} \rightarrow 0 \end{aligned}$$

Therefore $Y_m - Y_n \xrightarrow{\text{a.s.}} 0$ by part (b). ■

Egorov's Theorem. If μ is finite, then $X_n \xrightarrow{\text{a.s.}} X \supset X_n \xrightarrow{\text{a.u.}} X$.

Proof. If m is any positive integer, then $\lim_{n \rightarrow \infty} \mu \bigcup_{k=n}^{\infty} (|X_k - X| \geq \frac{1}{m}) = 0$

by the convergence a.e. criterion (a). Therefore, given $\epsilon > 0$, $n(m)$ can

be chosen so that $\mu A_m = \mu \bigcup_{k=n(m)}^{\infty} (|X_k - X| \geq \frac{1}{m}) < \frac{\epsilon}{2^n}$, for $m = 1, 2, \dots$.

Let $A = \bigcup A_m$. Then $\mu A \leq \sum \mu A_m < \sum \frac{\epsilon}{2^m} = \epsilon$. But

$$A^c = \bigcap A_m^c = \bigcap_{m=1}^{\infty} \bigcap_{k=n(m)}^{\infty} (|X_k - X| < \frac{1}{m}),$$

thus if $\omega \in A^c$ then, for $k \geq n(m)$, $|X_k(\omega) - X(\omega)| < \frac{1}{m}$ unif. on

$A^c \supset X_k \xrightarrow{u} X$ on A^c . ■

A measure μ is σ -finite iff there exist sets $A_n \in \mathcal{A}$ such that $\mu A_n < \infty$ and $\Omega = \sum A_n$.

Lusin's Theorem. If μ is σ -finite and $X_n \xrightarrow{\text{a.s.}} X$, then there is a

μ -null set N and a countable partition $N^c = \sum A_j$, where each $A_j \in \mathcal{A}$, and $\mu A_j < \infty$, such that $X_n \xrightarrow{u} X$ on A_j for each fixed j .

Proof. Since μ is σ -finite, it suffices to assume that μ is finite.

Using Egorov's Theorem, select A_1 so that $\mu A_1 < 1$ and $X_n \xrightarrow{u} X$ on A_1^c . For $n = 2$, select A_2 so that $\mu A_1 A_2 \leq \mu A_2 < \frac{1}{2}$ and $X_n \xrightarrow{u} X$ on A_2^c .

In general, select A_n so that $\mu \bigcap_{k=1}^n A_k \leq \mu A_n < \frac{1}{n}$ and $X_n \xrightarrow{u} X$ on A_n^c .

By the Continuity Theorem, $\mu \bigcap_{k=1}^{\infty} A_k = \lim \mu \bigcap_{k=1}^n A_k = 0$. Let $A = \bigcap_{k=1}^{\infty} A_k$.

Then $A^c = \bigcup_{k=1}^{\infty} A_k^c$, which can be written as a countable sum of sets, on each of which $X_n \xrightarrow{u} X$. Thus the theorem is true with $A = N$.

Diagonal Convergence Theorem.

(a) If μ is finite, $X_{mn} \xrightarrow{a.s.} X_m$ ($n \rightarrow \infty$), and $X_m \xrightarrow{a.s.} X$, then there exist sequences m_k, n_k of positive integers such that $X_{m_k n_k} \xrightarrow{a.s.} X$ ($k \rightarrow \infty$).

(b) If μ is σ -finite, say $\Omega = \sum A_j$ where $\mu A_j < \infty$ for each j , $X_{mn} \xrightarrow{a.s.} X_m$ ($n \rightarrow \infty$), and $X_m \xrightarrow{a.s.} X$, then for each j there exist sequences $m(j,k), n(j,k)$ of positive integers such that $X_{m(j,k)n(j,k)} \xrightarrow{a.s.} X$ ($k \rightarrow \infty$) on A_j , thus

$$X'_k = \sum_{j=1}^k I_{A_j} X_{m(j,k)n(j,k)} \xrightarrow{a.s.} X \quad (k \rightarrow \infty) \quad \text{on } \Omega.$$

If the X_{mn} are simple functions, then so are the X'_k .

Proof. Since $X_m \xrightarrow{a.s.} X$, for each $k = 1, 2, \dots$ there is a set $A_k \in \mathcal{A}$ such that $\mu A_k < 2^{-k}$ and $X_m \xrightarrow{u} X$ on A_k^c , using Egorov's theorem. Choose m_k so that $|X_{m_k} - X| < 2^{-k}$ uniformly on A_k^c . Now for each

k , $X_{m_k n_k} \xrightarrow{a.s.} X_{m_k}$ on A_k . Therefore, again by Egoroff's theorem, there is a set $B_k \subset A_k^c$ such that $\mu B_k < 2^{-k}$, and then n_k can be chosen so that $|X_{m_k n_k} - X_{m_k}| < 2^{-k}$ uniformly on $A_k^c B_k^c$.

Let

$$A = \bigcap_{k=1}^{\infty} \bigcup_{\ell=k}^{\infty} A_{\ell} \quad \text{and} \quad B = \bigcap_{i=1}^{\infty} \bigcup_{j=i}^{\infty} B_j.$$

Then $\mu A = \lim_{k \rightarrow \infty} \mu \bigcup_{j=k}^{\infty} A_j \leq \lim_{k \rightarrow \infty} \sum_{j=k}^{\infty} 2^{-j} = 0$, and similarly $\mu B = 0$.

$$A^c = \bigcup_{k=1}^{\infty} \bigcap_{\ell=k}^{\infty} A_{\ell}^c, \quad B^c = \bigcup_{i=1}^{\infty} \bigcap_{j=i}^{\infty} B_j^c.$$

Let $N = A \cup B$, so that

$$N^c = A^c B^c = \bigcup_{k=1}^{\infty} \bigcup_{i=1}^{\infty} \bigcap_{\ell=k}^{\infty} \bigcap_{j=i}^{\infty} A_{\ell}^c B_j^c.$$

Thus if $\omega \in N^c$, then for some q , $\omega \in A_k^c B_k^c$ for all $k \geq q$.

$$\begin{aligned} \therefore |X_{m_k n_k}(\omega) - X(\omega)| &\leq |X_{m_k n_k}(\omega) - X_{m_k}(\omega)| + |X_{m_k}(\omega) - X(\omega)| \\ &< 2^{-k} + 2^{-k} = 2^{-k+1} \end{aligned}$$

for all $k \geq q$, which means that $X_{m_k n_k} \rightarrow X$ on N^c . Part (b) follows directly from (a), using σ -finiteness. □

Let X be a Banach space with norm topology \mathcal{U} and Borel σ -field $\mathcal{B} = \overline{\mathcal{U}}$, and suppose μ is a σ -finite measure on \mathcal{A} .

A simple function means a function of the form $X = \sum_1^m x_j 1_{A_j}$,

where $A_j \in \mathcal{A}$ and $\mu A_j < \infty$. The class of all simple functions is again denoted by \mathcal{M}_s , and the integral is defined by $\int X = \sum_1^m x_j \mu A_j$, exactly as in the scalar case.

A function $X : \Omega \rightarrow X$ is said to be

(1) $(\mathcal{A}, \mathcal{B})$ -measurable iff $X^{-1}B \in \mathcal{A}$,

(2) strongly measurable iff there exists a sequence of simple functions X_n such that $X_n \xrightarrow{a.s.} X$,

(3) weakly measurable iff $fX : \Omega \rightarrow Z$ is a measurable scalar function for every continuous linear functional $f : X \rightarrow Z$.

These classes of functions will be denoted by $\mathcal{M}(\mathcal{A}, \mathcal{B})$, \mathcal{M} , and \mathcal{M}_w respectively; if $X = Z$, then they are clearly identical. It follows immediately that \mathcal{M} and \mathcal{M}_w are linear spaces, but $\mathcal{M}(\mathcal{A}, \mathcal{B})$ is not necessarily a linear space unless X is separable. In fact NEDOMA has given a counterexample in case the cardinal number of X is greater than the power of the continuum.

If $X \in \mathcal{M}$ and $g : X \rightarrow Z$ is continuous, it follows immediately that gX is a measurable scalar function, since any function of a simple function is simple. Therefore in particular $\mathcal{M} \subset \mathcal{M}_w$.

If $X_n \in \mathcal{M}$ and $X_n \xrightarrow{a.s.} X$, then it follows from the Diagonal Convergence Theorem that $X \in \mathcal{M}$, since μ is σ -finite. If $X_n \in \mathcal{M}_w$ and $X_n \xrightarrow{a.s.} X$, then it follows immediately that $X \in \mathcal{M}_w$, since the class of all measurable scalar functions is closed under limits.

Measurable Functions Theorem. If X is separable, then every $(\mathcal{A}, \mathcal{B})$ -measurable function $X : \Omega \rightarrow X$ is strongly measurable, thus

$$\mathcal{M}(\mathcal{A}, \mathcal{B}) \subset \mathcal{M} \subset \mathcal{M}_w.$$

Proof. Let $[v_n]$ be a countable dense set, $\varepsilon > 0$, and $S_n = S(x_n, \varepsilon)$.

Then $X = \bigcup S_n = \sum B_n$, where $B_n = S_n \cdot \bigcap_{k=1}^{n-1} S_k^c \in \mathcal{B}$; hence

$\Omega = \sum X^{-1} B_n = \sum A_n$, where $A_n = X^{-1} B_n \in \mathcal{A}$ since $X \in \mathcal{M}(\mathcal{A}, \mathcal{B})$.

Define $X_\varepsilon(\omega) = \sum v_n I_{A_n}(\omega)$, and let any $\omega \in \Omega$ be given, say $\omega \in A_n = X^{-1} B_n$. Then $X(\omega) \in B_n \subset S_n$, hence $|X_\varepsilon(\omega) - X(\omega)| = |v_n - X(\omega)| < \varepsilon$.

This implies that there exists a sequence X_n of elementary functions such that $X_n \xrightarrow{u} X$ everywhere. Since μ is σ -finite, it now follows from the Diagonal Convergence Theorem that there is a sequence X'_n of simple functions such that $X'_n \xrightarrow{a.s.} X$.

If X is weakly measurable and \mathcal{J} is the usual topology on Z , then $(fX)^{-1}\mathcal{J} = X^{-1}(f^{-1}\mathcal{J}) \subset \mathcal{A}$ for every continuous linear functional $f: X \rightarrow Z$; and $\mathcal{J} = \bigcup_f f^{-1}\mathcal{J} \subset \mathcal{U}$ since each f is continuous. In the next chapter it will be shown, with the aid of the Hahn-Banach Theorem, that $\bar{\mathcal{U}} = \bar{\mathcal{J}}$ whenever X is separable. But recall that $X^{-1}\mathcal{J} \subset \mathcal{A}$ implies $X^{-1}\bar{\mathcal{J}} \subset \mathcal{A}$. Therefore every weakly measurable function is $(\mathcal{A}, \mathcal{B})$ -measurable in case X is separable, which means that all three notions of measurability are the same. □

A function $X: \Omega \rightarrow X$ is said to be a.s. separably valued iff there exists a μ -null set N such that $X[N^c]$ is separable. It is a rather remarkable fact, proved by PETTIS, that every strongly measurable function is a.s. separably valued. Because of this and other considerations, including the example of NEDOMA mentioned above, X is usually assumed to be separable.

It is easily seen that $|\int X| \leq \int |X|$ for any $X \in \mathcal{M}_s$. A norm is defined on the linear space \mathcal{M}_s by writing

$$\|X\| = \int |X| = \int |X| \, d\mu.$$

The Bochner integral of a strongly measurable function is then defined by completing the normed linear subspace \mathcal{M}_S with respect to this norm, using the Cantor completion method.

Definition. A function $X \in \mathcal{M}$ is Bochner summable iff there exists a sequence $X_n \in \mathcal{M}_S$ such that $X_n \xrightarrow{a.s.} X$ and

$$\|X_n - X\| = \int |X_n - X| \rightarrow 0.$$

It follows that $\int X_n$ is a Cauchy sequence, and by definition

$$\int X = \lim \int X_n.$$

If $Y_n \in \mathcal{M}_S$ is another such sequence, then $|\int X_m - \int Y_n| \leq \int |X_m - Y_n|$ implies that $\lim \int X_m = \lim \int Y_n$, so $\int X$

is well defined. Since every Cauchy sequence is bounded, it follows easily that $|\int X| \leq \int |X| = \|X\| < \infty$.

This method is due to DUNFORD [1]. Actually he starts out slightly further down the line, with Cauchy sequences in \mathcal{M}_S . But it is not difficult to show, using the Markov inequality

$$[|X_m - X_n| \geq \epsilon] \leq \frac{1}{\epsilon} \|X_m - X_n\|$$

and the Convergence Theorem, that every Cauchy sequence $X_n \in \mathcal{M}_S$ has a subsequence $X'_n \xrightarrow{a.s.} X \in \mathcal{M}$.

The class of all Bochner summable functions will be denoted by \mathcal{A} , just as in the scalar case. It follows directly from the definition

that \mathcal{L} is a NLS with norm $\|X\| = \int |X|$. In case $X = Z$, it is obvious from the definitions that a function $X : \Omega \rightarrow Z$ is Bochner summable iff it is Lebesgue summable in the old sense, so that the definition of the Bochner integral is consistent.

A function $X : \Omega \rightarrow X$ is said to be square summable iff there is a sequence $X_n \in \mathcal{M}_S$ such that $X_n \xrightarrow{a.s.} X$ and

$$\|X_n - X\|_2^2 = \int |X_n - X|^2 \rightarrow 0.$$

It follows from the Schwarz inequality (for real functions) that

$$\|X + Y\|_2 \leq \|X\|_2 + \|Y\|_2, \text{ whence } \|X_n + Y_n - X - Y\|_2 \leq \|X_n - X\|_2 + \|Y_n - Y\|_2.$$

Therefore the class \mathcal{L}^2 of square summable functions is a NLS with

$$\|X\|_2^2 = \int |X|^2 < \infty.$$

Notice that a square summable function is not necessarily summable. But it follows from the Schwarz inequality that for any subset $A \in \mathcal{A}$ such that $\mu A < \infty$,

$$\int_A |X_n - X| \leq (\mu A) \left[\int_A |X_n - X|^2 \right]^{\frac{1}{2}}.$$

Hence every square summable function is summable over sets of finite measure. In particular, $\mathcal{L}^2 \subset \mathcal{L}$ if μ is finite.

Riesz-Fischer Theorem. The space \mathcal{L}^r ($r = 1, 2$) is equal to

$[X \in \mathcal{M} : \int |X|^r = \|X\|^r < \infty]$, and is a Banach space if functions are identified with their μ -equivalence sets.

Proof. It has already been noted that $X \in \mathcal{L}^r$ implies $\int |X|^r < \infty$, and that \mathcal{L}^r is a NLS with norm defined by $\|X\| = \left(\int |X|^r \right)^{1/r}$. Clearly

$\|X\| = 0$ implies $X \stackrel{a.s.}{=} 0$.

Now suppose $X \in \mathcal{M}$, $\int |X|^r < \infty$, and choose $X_n \in \mathcal{M}_s$ so that $X_n \stackrel{a.s.}{\rightarrow} X$. Since μ is σ -finite, by Lusin's Theorem there is a null set N and a countable partition $N^c = \sum A_k$ such that $A_k \in \mathcal{C}$, $0 < \mu A_k < \infty$, and $X_n \xrightarrow{u} X$ on A_k for each $k = 1, 2, \dots$. Let $\varepsilon > 0$ be given, and for each k choose n_k so that $|X_{n_k} - X|^r < \frac{\varepsilon}{2^k \mu A_k}$ uniformly on A_k ; then define $X_\varepsilon = \sum X_{n_k} 1_{A_k}$ and $X_{\varepsilon m} = \sum_{k=1}^m X_{n_k} 1_{A_k}$.

It follows by the Monotone Convergence Theorem that

$$\int |X_\varepsilon - X|^r = \sum \int_{A_k} |X_\varepsilon - X|^r \leq \sum \frac{\varepsilon}{2^k} = \varepsilon$$

and

$$\int |X_\varepsilon|^r = \sum \int_{A_k} |X_{n_k}|^r \leq 2 \int |X_\varepsilon - X|^r + 2 \int |X|^r < \infty,$$

hence

$$\int |X_\varepsilon - X_{\varepsilon m}|^r = \sum_{k=m+1} \int_{A_k} |X_{n_k}|^r \rightarrow 0 \text{ as } m \rightarrow \infty.$$

Note that X_ε is an elementary function and $X_{\varepsilon m}$ is a simple function for each m , since $\mu A_k < \infty$. If m is now chosen so that $\int |X_\varepsilon - X_{\varepsilon m}| < \varepsilon$, then

$$\int |X - X_{\varepsilon m}|^r \leq 2 \int |X - X_\varepsilon|^r + 2 \int |X_\varepsilon - X_{\varepsilon m}|^r < 4\varepsilon.$$

Thus if $X \in \mathcal{M}$ and $\int |X|^r < \infty$, then there is a sequence $X_n' \in \mathcal{M}_s$ such that $\int |X_n' - X|^r = \|X_n' - X\|^r \rightarrow 0$. But this implies that

$$\mu[|X_n^i - X| \geq \varepsilon] \leq \varepsilon^{-r} \|X_n^i - X\|^r \rightarrow 0$$

for every $\varepsilon > 0$ by the Markov-Tchebychev inequality. It then follows from the Convergence theorem that there is a subsequence X_n'' such that $X_n'' \xrightarrow{\text{a.s.}} X$; and clearly $\int |X_n'' - X|^r \rightarrow 0$. Therefore

$$\mathcal{L}^r = [X \in \mathcal{M} : \int |X|^r < \infty] .$$

Now suppose X_n is a Cauchy sequence in \mathcal{L}^r . Then $\mu[|X_m - X_n| \geq \varepsilon] \leq \varepsilon^{-r} \|X_m - X_n\|^r \rightarrow 0$ for every $\varepsilon > 0$ by the Markov-Tchebychev inequality, so by the Convergence Theorem there is a subsequence Y_n such that $Y_m - Y_n \xrightarrow{\text{a.s.}} 0$. Since \mathcal{X} is complete, $Y_n \xrightarrow{\text{a.s.}} Y$; and $Y \in \mathcal{M}$ since \mathcal{M} is closed under a.s. limits.

Now $X_m - Y_n \xrightarrow{\text{a.s.}} X_m - Y$ ($n \rightarrow \infty$) for each m , so by Fatou's lemma it follows that

$$\int |X_m - Y|^r = \int \lim_n |X_m - Y_n|^r \leq \lim_n \int |X_m - Y_n|^r \rightarrow 0 \quad (m \rightarrow \infty)$$

and $\int |Y|^r \leq \lim_n \int |Y_n|^r < \infty$ since every Cauchy sequence is bounded.

This means \mathcal{L}^r is complete, hence is a Banach space, which completes the proof. ■

Dominated Convergence Theorem. If $X_n \in \mathcal{L}$, $X_n \xrightarrow{\text{a.s.}} X$, and $|X_n| \leq g \in \mathcal{L}_{\text{real}}$, then $X \in \mathcal{L}$ and $\int X_n \rightarrow \int X$.

Proof. Clearly $|X| = \lim |X_n| \leq g$ a.s.; hence $\int |X| < \infty$ and the previous theorem implies $X \in \mathcal{L}$. Furthermore $|X_n - X| \rightarrow 0$, $|X_n - X| \leq 2g$

implies $\int |X_n - X| \rightarrow 0$ by the Lebesgue DCT, and

$$|\int X_n - \int X| \leq \int |X_n - X| \rightarrow 0 \triangleright \int X_n \rightarrow \int X.$$

Hilbert Space. Now consider the space \mathcal{L}^2 in the case $X = \mathbb{Z}$. Note that $X, Y \in \mathcal{L}^2$ implies $XY^* \in \mathcal{L}^2$ by the Schwarz inequality. Define $X \circ Y = \int XY$, called the inner product (dot product) of X and Y , and note that $\|X\|^2 = X \circ X$. The space \mathcal{L}^2 is an important example of a Hilbert space. Since Hilbert spaces occur frequently and have many of the properties of Euclidean space that Banach spaces do not have in general, it is appropriate to develop some of their basic properties here. For a more complete discussion see AKHIESER and GLAZMAN, HALMOS [2], KOLMOGOROV and FOMIN, RIESZ and SZ. NAGY, SIMMONS, or VON NEUMANN.

Let X be a linear space. An inner product on X is a scalar valued function \circ on $X \times X$, with $\circ(x, y)$ denoted by $x \circ y$, such that

$$(1) \quad (x + y) \circ z = x \circ z + y \circ z$$

$$(2) \quad (\alpha x) \circ y = \alpha(x \circ y)$$

$$(3) \quad x \circ y = (y \circ x)^*$$

$$(4) \quad x \circ x > 0 \text{ if } x \neq \theta$$

An inner product space (IPS) is a pair (X, \circ) consisting of a linear space and an inner product on it. It follows that $\theta \circ y = x \circ \theta = 0$, $x \circ (y + z) = x \circ y + x \circ z$, and $x \circ (\alpha y) = \alpha(x \circ y)$. Two vectors x, y are said to be orthogonal, written $x \perp y$, iff $x \circ y = 0$. If $M \subset X$, then $x \perp M$ means $x \perp y$ for all $y \in M$, and $M^\perp = [x : x \perp M]$.

By definition $|x| = (x \circ x)^{1/2}$, and it follows that $|\alpha x| = |\alpha| |x|$, $|x| = 0 \triangleright x = \theta$, and

$$|x + y|^2 = |x|^2 + 2\operatorname{Re}(x \cdot y) + |y|^2 \leq |x|^2 + 2|x \cdot y| + |y|^2.$$

Thus the triangle inequality follows from

$$|x \cdot y| \leq |x||y| \quad (\text{Schwarz Inequality}).$$

If u and v are unit vectors such that $u \cdot v > 0$, then the Schwarz inequality follows immediately from $|u - v|^2 = 2 - 2|u \cdot v| \geq 0$, and it is trivially true for $x \perp y$. Otherwise let $u = (e^{-i\alpha}) \frac{x}{|x|}$ and $v = \frac{y}{|y|}$, where $\alpha = \arg x \cdot y$. The equality sign holds in the Schwarz inequality if and only if x and y are proportional; hence the equality sign holds in the triangle inequality if and only if x and y are proportional with positive proportionality constant. A Hilbert space is an IPS which is also a Banach space with respect to the norm defined by $|x|^2 = x \cdot x$.

It follows by expansion that $|\sum_1^n x_j|^2 = \sum_1^n \sum_1^n x_j \cdot x_k$. Hence if the vectors x_j are orthogonal, then

$$|\sum_1^n x_j|^2 = \sum_1^n |x_j|^2 \quad (\text{Pythagorean Law}).$$

A sequence $[y_n]$ is orthonormal iff $y_j \cdot y_k = \delta_{jk}$. Suppose $[y_n]$ is an orthonormal sequence, and let x be an arbitrary vector. If

$$x_n = \sum_1^n (x \cdot y_k) y_k$$

then

$$x \cdot x_n = x_n \cdot x_n = |x_n|^2 = \sum_1^n |x \cdot y_k|^2,$$

hence $(x - x_n) \cdot x_n = 0$ and

$$|x|^2 = |x - x_n|^2 + |x_n|^2 \geq \sum_1^n |x - y_k|^2$$

by the Pythagorean Law. It follows that

$$\sum |x \cdot y_n|^2 \leq |x|^2 \quad (\text{Bessel's Inequality})$$

which implies that $\sum (x \cdot y_n) y_n$ always converges, since the partial sums are a Cauchy sequence and X is complete. But clearly $x_n \rightarrow x$ if and only if $|x_n|^2 \rightarrow |x|^2$, so the equality sign holds in Bessel's inequality if and only if $\sum (x \cdot y_n) y_n = x$.

If X is a Hilbert space, then a set $[y_i : i \in I]$ of distinct orthonormal vectors in X is called a basis for X iff it is a maximal set in the class \mathcal{O} of all orthonormal sets:

$$[y_i : i \in I] \subset [z_j : j \in J] \text{ } \mathcal{O} \triangleright [y_i : i \in I] = [z_j : j \in J].$$

It follows that $[y_i : i \in I]$ is a basis for X if and only if $x \cdot y_i$ for all $i \in I$ implies $x = \theta$.

Assume $X \neq [\theta]$. Then \mathcal{O} is nonempty and partially ordered by inclusion, and every chain $\zeta \subset \mathcal{O}$ is easily seen to have an upper bound, namely $\bigcup \zeta$. Therefore, by Zorn's lemma, there always exists a basis.

If X is separable, then every basis for X is countable. For suppose $[x_n]$ is a countable dense set, and let $[y_i]$ be a basis. If $i \neq j$, then $|y_i - y_j|^2 = |y_i|^2 + |y_j|^2 = 2$. Thus if $[y_i]$ is uncountable, then every dense set must be uncountable since the spheres $S(y_i, 1)$ are disjoint.

Basis Theorem. A countable orthonormal set $[y_n]$ is a basis for X iff $x = \sum (x \cdot y_n) y_n$ for all $x \in X$.

Proof. By Bessel's inequality it follows that $\sum (x \cdot y_n) y_n$ converges to some vector, say y . But $y \cdot y_m = x \cdot x_m \triangleright (y - x) \cdot y_m \triangleright y = x$, by completeness.

Conversely, if $\sum (x \cdot y_n) y_n = x$ for all x , then $(x \cdot y_n) = 0$ for all n obviously implies $x = \theta$. □

In particular, if $[y_n]$ is a countable basis, then X is necessarily separable, since the set $[z_n]$ of all finite linear combinations of the vectors y_n , with rational coefficients, is dense in X . Thus X is separable if and only if it has a countable basis.

It follows by expansion that

$$|x + y|^2 + |x - y|^2 = 2|x|^2 + 2|y|^2 \quad (\text{Parallelogram Law})$$

$$\text{and } 4(x \cdot y) = |x + y|^2 - |x - y|^2 + i|x + iy|^2 - i|x - iy|^2.$$

The parallelogram law is characteristic of Hilbert space. In fact, VON NEUMANN showed that if X is a NLS satisfying the parallelogram law, then the above expression defines an inner product.

A subset M of a linear space is a manifold iff $x, y \in M$ implies $\alpha x + \beta y \in M$. The following theorem shows that Hilbert space has one of the most important properties of Euclidean space.

Perpendicular Distance Theorem. If M is a subspace (closed manifold) of a Hilbert space X and $x \in X$, then there is a unique vector $x_0 \in M$ such that $|x - x_0| = \inf [|x - y| : y \in M] = d$. Furthermore $x - x_0 = z \perp M$.

Proof. Choose $[x_n] \subset M$ so that $|x - x_n| \rightarrow d$, and note that by the Parallelogram Law,

$$|(x - x_m) + (x - x_n)|^2 + |(x - x_m) - (x - x_n)|^2 = 2|x - x_m|^2 + 2|x - x_n|^2,$$

hence $|x_n - x_m|^2 = 2|x - x_m|^2 + 2|x - x_n|^2 - 4|x - \frac{1}{2}(x_m + x_n)|^2$. But $\frac{1}{2}(x_m + x_n) \in M$ implies $d \leq |x - \frac{1}{2}(x_m + x_n)| \leq \frac{1}{2}|x - x_m| + \frac{1}{2}|x - x_n| \rightarrow d$ as $m, n \rightarrow \infty$, hence $|x_n - x_m|^2 \rightarrow 0$ and $[x_n]$ is a Cauchy sequence.

Let $x_0 = \lim x_n$. Then $x_0 \in M$ since M is closed, and $d \leq |x - x_0| \leq |x - x_n| + |x_n - x_0| \rightarrow d$ implies $|x - x_0| = d$. If $x_0, x_1 \in M$ and $|x - x_0| = |x - x_1| = d$, then $\frac{1}{2}(x_0 + x_1) \in M$ and $|x_0 - x_1|^2 = 2|x - x_0|^2 + 2|x - x_1|^2 - 4|x - \frac{1}{2}(x_0 + x_1)|^2 \leq 2d^2 + 2d^2 - 4d^2 = 0$, hence $|x_0 - x_1| = 0$ and $x_0 = x_1$.

Now suppose $x - x_0 = z$ is not \perp to every vector in M . Then there is a unit vector $u \in M$ such that $u \cdot z = \alpha \neq 0$ and it suffices to assume that $\alpha > 0$ (otherwise use $ue^{-i \arg \alpha}$). Now $y = x_0 + \alpha u \in M$, but $|x - y|^2 = |z - \alpha u|^2 = |z|^2 - \alpha^2 = d^2 - \alpha^2 < d^2$, which is a contradiction. □

F. Riesz Representation Theorem. If $f : X \rightarrow \mathbb{C}$ is a continuous linear functional on a Hilbert space X , then there exists a unique vector $y \in X$ such that $f(x) = x \cdot y$ for all $x \in X$. Furthermore $|y| = \sup_{|x|=1} |fx|$.

Proof. Let $M = [x : f(x) = 0]$. Then M is a subspace since f is continuous. If $M = X$, take $y = 0$. Otherwise, by the above theorem there exists a unit vector $u \in M^\perp$, and $M_u = [(fx)u - (fu)x : x \in X] \subset M$. Therefore, $[(fx)u - (fu)x] \cdot u = f(x) - (fu)x \cdot u = 0$, and it suffices to let $y = (fu)^*u$. If $fx = x \cdot y_1 = x \cdot y_2$, then $x \cdot (y_1 - y_2) = 0$ for all x , hence $y_1 = y_2$.

Now $|fx| = |x \cdot y| \leq |x||y|$ by the Schwarz inequality, hence

$\sup_{|x|=1} |fx| \leq |y|$. But if $y \neq \theta$, then $\frac{y}{|y|} = u$ is a unit vector and

$fu = |y|$, hence $\sup |fx| = |y|$ if $y \neq \theta$. If $y = \theta$, this is trivially true.

This beautiful theorem, which completely characterizes all continuous linear functionals on Hilbert space, has many important applications. For example, suppose $X : \Omega \rightarrow X$ is weakly measurable, where $(\Omega, \mathcal{A}, \mu)$ is a measure space, σ -finite or not, and X is a Hilbert space. This means $X \circ y$, hence also $y \circ X$, is a measurable scalar function for every $y \in X$. If there exists a scalar function $\alpha(\omega)$ such that $|X(\omega)| \leq \alpha(\omega) \in \mathcal{L}$, then

$$g(y) = \int y \circ X(\omega) \, d\mu(\omega)$$

exists and is a continuous linear functional on X . Therefore, by the F. Riesz Representation Theorem, there is a unique vector $z \in X$ such that $g(y) = y \circ z$, and by definition $z = \int X \, d\mu$. This is called the Pettis integral, or weak integral, and its defining equation is

$$y \circ \int X = \int y \circ X, \quad y \in X.$$

The Pettis integral may exist even when the Bochner integral does not. But it is easy to see that every Bochner summable function is Pettis summable; and the two integrals are equal since the defining equation of the Pettis integral holds for simple functions, hence also Bochner summable functions.

Suppose now that X is separable, and let $[y_n]$ be a basis. If $(\Omega, \mathcal{A}, \mu)$ is a σ -finite measure space, then it follows that every weakly

measurable function $X : \Omega \rightarrow X$ is strongly measurable, since

$$X = \sum (X \circ \gamma_n) \gamma_n$$

and every partial sum is strongly measurable. Thus in this case a function is Bochner summable if and only if it is Pettis summable. Furthermore,

$\int_A X \equiv \theta \quad [A \in \mathcal{Q}]$ implies by the scalar case that $X \circ \gamma_n \stackrel{a.s.}{=} 0$. Thus

$$X = \sum (X \circ \gamma_n) \gamma_n \stackrel{a.s.}{=} 0$$

since the union of a countable number of μ -null sets is μ -null. This will be shown in the next chapter for separable Banach spaces, using the Hahn-Banach Theorem.

Finally, consider the space \mathcal{L}^2 of all square summable functions $X : \Omega \rightarrow X$, and for $X, Y \in \mathcal{L}^2$ define

$$X : Y = \int X \cdot Y.$$

This is clearly an inner product, and $X : X = \|X\|^2$; hence an \mathcal{L}^2 space of Hilbert valued functions is again a Hilbert space, by the Riesz-Fischer Theorem.

CHAPTER IV

PROBABILITY THEORY

In this chapter, the material developed so far will be used to generalize some theorems of classical probability theory. Throughout the chapter, (Ω, \mathcal{A}, P) is an arbitrary probability space, and (X, \mathcal{B}) is a measurable space; usually X is a separable Hilbert or Banach space with norm topology \mathcal{U} and Borel σ -field $\mathcal{B} = \bar{\mathcal{U}}$.

A random variable (r.v.) in X (X -valued r.v.) means a function $X : \Omega \rightarrow X$ such that $X^{-1}\mathcal{B} \subset \mathcal{A}$, and will be called a Hilbert r.v. or Banach r.v. according as X is a Hilbert or Banach space. The expectation (expt.) of a Banach r.v. X is defined as its Bochner integral whenever it exists, and is denoted by either EX or $\int X$ according to convenience. The pr. distribution of a r.v. X is the pr. Q on \mathcal{B} defined by

$$QB = P[X \in B] = PX^{-1}B \quad (B \in \mathcal{B}),$$

and is usually denoted by P_X .

Most of the theorems in this chapter are generalizations of theorems in LOEVE and in HALMOS [1], for real r.v.'s, to the case of Hilbert or Banach r.v.'s. The results concerning conditional expectations are mostly due to DRIMLAND and HANS. For a more complete discussion of recent developments, see GRENANDER.

A class \mathcal{M} of subsets of X is a monotone class iff, for every monotone sequence of sets $A_n \in \mathcal{M}$, $\lim A_n \in \mathcal{M}$.

If \mathcal{C} is an arbitrary class of subsets of X , then by definition

$$\bar{\mathcal{C}} = [\mathcal{B} : \mathcal{C} \subset \mathcal{B}, \mathcal{B} \text{ is a } \sigma\text{-field}]$$

$$\hat{\mathcal{C}} = [\mathcal{F} : \mathcal{C} \subset \mathcal{F}, \mathcal{F} \text{ is a field}]$$

$$\hat{\mathcal{C}} = [\mathcal{M} : \mathcal{C} \subset \mathcal{M}, \mathcal{M} \text{ is a monotone class}]$$

It follows easily that $\bar{\mathcal{C}} = \hat{\bar{\mathcal{C}}}$. Similarly, \mathcal{C}^r denotes the minimal ring containing \mathcal{C} , and \mathcal{C}^s denotes the minimal σ -ring containing \mathcal{C} . A field \mathcal{C} is a σ -field iff it is a monotone class, since $\bigcup_n A_n = \bigcup_n \bigcup_{k=1}^n A_k$. Thus it is clear that $\hat{\mathcal{C}} \subset \hat{\bar{\mathcal{C}}} = \bar{\mathcal{C}}$ for any class \mathcal{C} ; and if $\hat{\mathcal{C}}$ is a field, then also $\bar{\mathcal{C}} \subset \hat{\bar{\mathcal{C}}} = \hat{\mathcal{C}}$, hence $\bar{\mathcal{C}} = \hat{\mathcal{C}}$. These remarks, and the following theorem, remain true for rings, σ -rings, and \mathcal{C}^s in place of fields, σ -fields, and $\bar{\mathcal{C}}$.

Monotone Class Theorem. If \mathcal{C} is a field, so is $\hat{\mathcal{C}}$, and

$$\bar{\mathcal{C}} = \hat{\mathcal{C}}.$$

Proof. For each $B \in \hat{\mathcal{C}}$, let $\hat{\mathcal{C}}_B = [A \in \hat{\mathcal{C}} : A \cup B, A^c B, AB^c \in \hat{\mathcal{C}}]$. It follows that $\hat{\mathcal{C}}_B$ is a monotone class. If $B \in \mathcal{C}$, then $\mathcal{C} \subset \hat{\mathcal{C}}_B$, hence $\hat{\mathcal{C}} \subset \hat{\mathcal{C}}_B$. Thus if $A \in \hat{\mathcal{C}}$ is fixed, then $A \in \hat{\mathcal{C}}_B$ for each $B \in \mathcal{C}$. This implies that $\mathcal{C} \subset \hat{\mathcal{C}}_A$, hence $\hat{\mathcal{C}} \subset \hat{\mathcal{C}}_A$. Therefore $\hat{\mathcal{C}} = \hat{\mathcal{C}}_A$ for each $A \in \hat{\mathcal{C}}$, which implies that $\hat{\mathcal{C}}$ is a field. ■

One of the most important concepts in pr. theory is the concept of independence. Let the index t vary over an arbitrary index set T , and suppose $\mathcal{C}_t \subset \mathcal{A}$ are given classes of events. The classes \mathcal{C}_t are said to be independent (ind.) with respect to the pr. P iff, for every finite subset $[t_1, t_2, \dots, t_n] \subset T$ of distinct indices and every choice

of events $A_j \in \mathcal{Z}_{t_j}$,

$$P \bigcap_{j=1}^n A_j = \prod_{j=1}^n P A_j.$$

Random variables $X_t : \Omega \rightarrow X$ are independent iff the σ -fields $X_t^{-1}\mathcal{B}$ are independent.

Independence Theorem

If \mathcal{Z}_t are independent classes closed under finite intersections, then $\bar{\mathcal{Z}}_t$ are independent. Random variables X_t in a measurable space $(X, \bar{\mathcal{E}})$, where \mathcal{E} is closed under finite intersections, are independent iff $\mathcal{Z}_t = X_t^{-1}\mathcal{E}$ are independent.

Proof. Assume without loss of generality that $T = [1, 2, \dots, n]$, and that $\emptyset, \Omega \in \mathcal{Z}_k$. Consider first the case $n = 2$, and let

$$\mathcal{M}_1 = [A_1 \in \bar{\mathcal{Z}}_1 : P A_1 A_2 = P A_1 \cdot P A_2 \text{ for all } A_2 \in \mathcal{Z}_2].$$

By hypothesis $\mathcal{Z}_1 \subset \mathcal{M}_1$, and it follows easily that \mathcal{M}_1 is closed under finite sums and proper differences; but it is not obvious that \mathcal{M}_1 is closed under finite intersections.

Let \mathcal{Z}'_1 be the class of all finite unions of \mathcal{Z}_1 sets. Then \mathcal{Z}'_1 is closed under finite unions and intersections; and it follows by induction that $\mathcal{Z}'_1 \subset \mathcal{M}_1$, using the fact that \mathcal{Z}_1 is closed under finite intersections.

Let \mathcal{D}_1 be the class of all proper differences of \mathcal{Z}'_1 sets, and let \mathcal{F}_1 be the class of all finite sums of \mathcal{D}_1 sets. Then $\mathcal{F}_1 \subset \mathcal{M}_1$, and \mathcal{F}_1 is a field (see Chapter VI); in fact $\mathcal{F}_1 = \mathcal{Z}'_1$. Finally, \mathcal{M}_1 is

a monotone class, hence $\mathcal{F}_1 \subset \mathcal{M}_1$ implies $\bar{\mathcal{F}}_1 = \bar{\mathcal{F}}_1 = \hat{\mathcal{F}}_1 \subset \mathcal{M}_1$ by the Monotone Class theorem. Therefore $\bar{\mathcal{F}}_1$ and \mathcal{Z}_2 are independent. Since $\bar{\mathcal{F}}_1$ is closed under finite intersections, it follows in the same way that $\bar{\mathcal{F}}_1$ and $\bar{\mathcal{Z}}_2$ are independent.

Now suppose $\bar{\mathcal{F}}_1, \dots, \bar{\mathcal{F}}_{n-1}$ are independent, and let

$$\mathcal{E}_{n-1} = \left[\bigcap_{k=1}^{n-1} A_k : A_k \in \mathcal{Z}_k \right].$$

Then \mathcal{E}_{n-1} and \mathcal{Z}_n are independent and closed under finite intersections, hence $\bar{\mathcal{E}}_{n-1}$ and $\bar{\mathcal{Z}}_n$ are independent. But

$$\bar{\mathcal{E}}_{n-1} = \left[\bigcup_{j=1}^{n-1} \mathcal{Z}_j \right]^- = \left[\bigcup_{j=1}^{n-1} \bar{\mathcal{Z}}_j \right]^-;$$

thus if $A_j \in \bar{\mathcal{Z}}_j$ ($j = 1, \dots, n$), then $A_1 \dots A_{n-1} \in \bar{\mathcal{E}}_{n-1}$ and

$P(A_1 \dots A_{n-1})A_n = P(A_1 \dots A_{n-1})PA_n = (PA_1 \dots PA_{n-1})PA_n$ using the induction hypothesis. This completes the proof of the first assertion, and the second assertion follows from the first, using the definitions. ■

Let (X, \mathcal{B}) be a measurable space, $X^n = \prod_{j=1}^n X_j$,

$$0 = n_0 < n_1 < \dots < n_m = n, \quad \Delta n_i = n_i - n_{i-1}, \quad m_i = n_{i-1} + 1,$$

and $X^{\Delta n_i} = \prod_{j=m_i}^{n_i} X_j$ ($i = 1, \dots, m$), where $X_j \equiv X$. The projections

$q_i : X^n \rightarrow X^{\Delta n_i}$ are defined by $q_i(x_1, \dots, x_n) = (x_{m_i}, \dots, x_{n_i})$, $\mathcal{Z}^{\Delta n_i}$ is the class of all "rectangles" in $X^{\Delta n_i}$ of the form $\prod_{j=m_i}^{n_i} B_j (B_j \in \mathcal{B})$,

$\mathcal{B}^{\Delta n_i} = \overline{\mathcal{Z}^{\Delta n_i}}$, and $p_j(x_1, \dots, x_n) = x_j$. It follows that

$$q_i^{-1} \mathcal{B}^{\Delta n_i} = \left[q_i^{-1} \mathcal{Z}^{\Delta n_i} \right]^- = \left[\bigcup_{j=m_i}^{n_i} p_j^{-1} \mathcal{B} \right]^-.$$

A product probability on the product σ -field $\mathcal{B}^n = \left[\bigcup_{j=1}^n \mathcal{P}_j^{-1} \mathcal{B} \right]^{-}$

means a pr. Q such that, if $B_j \in \mathcal{B}$ for $j = 1, \dots, n$, then

$$Q \prod_{j=1}^n B_j = \prod_{j=1}^n Q \mathcal{P}_j^{-1} B_j. \text{ The classes } q_i^{-1} \zeta^{\Delta n_i} \subset \mathcal{B}^n \text{ are closed under}$$

finite intersections and independent with respect to any product pr. Q

on \mathcal{B}^n . Therefore, by the Independence Theorem, the σ -fields

$$q_i^{-1} \mathcal{B}^{\Delta n_i} \subset \mathcal{B}^n \text{ are independent with respect to any product pr. In}$$

particular, suppose (Y, ζ) is another measurable space, let

$$f_i : \mathcal{X}^{\Delta n_i} \rightarrow Y \text{ be Borel functions, and define the Borel functions}$$

$$g_i : \mathcal{X}^n \rightarrow Y \text{ by } g_i = f_i q_i. \text{ Then } g_i^{-1} \zeta \subset q_i^{-1} \mathcal{B}^{\Delta n_i}, \text{ hence } g_i \text{ are}$$

independent with respect to any product pr. Q on \mathcal{B}^n .

Borel Functions Theorem

If $X_j : \Omega \rightarrow \mathcal{X}$ are independent r.v.'s ($j = 1, \dots, n$) and

$f_i : \mathcal{X}^{\Delta n_i} \rightarrow Y$ are Borel functions ($i = 1, \dots, m$), then

$Y_i = f_i(X_{m_i}, \dots, X_{n_i}) : \Omega \rightarrow Y$ are also independent.

Proof. Let $X = (X_j : j = 1, \dots, n) : \Omega \rightarrow \mathcal{X}^n$ be the vector function

with component functions X_j , and let P_X be the pr. distribution of

X . Then P_X is a product pr. on \mathcal{B}^n , for if $B_j \in \mathcal{B}$, then

$$P_X \prod_{j=1}^n B_j = P_X^{-1} \prod_{j=1}^n B_j = P \bigcap_{j=1}^n X_j^{-1} B_j = \prod_{j=1}^n P X_j^{-1} B_j = \prod_{j=1}^n P_X \mathcal{P}_j^{-1} B_j, \text{ since } X_j$$

are independent.

Using the previous notation, $Y_i = g_i X$; thus if $C_i \in \zeta$, then

$$P \bigcap_{i=1}^m Y_i^{-1} C_i = P_X \bigcap_{i=1}^m g_i^{-1} C_i = \prod_{i=1}^m P_X g_i^{-1} C_i = \prod_{i=1}^m P Y_i^{-1} C_i$$

since g_i are independent with respect to the product pr. P_X . ■

Change of Variables Theorem

Suppose (X, \mathcal{B}) is a measurable space and Y is a separable Banach space with norm topology \mathcal{U} and Borel σ -field $\mathcal{C} = \bar{\mathcal{U}}$, and let $X : \Omega \rightarrow X$, $f : X \rightarrow Y$ be r.v.'s. Then

$$\int_{\Omega} fX(\omega) dP = \int_X f(x) dP_X$$

whenever either side exists.

Proof. Suppose first that $f(x) = yIB$, where $y \in Y$ and $B \in \mathcal{B}$. Then $fX(\omega) = yIB[X(\omega)] = yIX^{-1}B$, and asserted equation becomes $yPX^{-1}B = yP_X B$, which is true by definition of P_X ; it follows easily that the integrals are equal whenever f is a simple function. If $Y = \mathbb{R}$ and $f \in \mathcal{M}^+$, the conclusion follows from the MCT; thus in any case the integrals are equal when the integrands are replaced by their norms, so that the existence of either side implies that of the other. But if the integral on the right exists, then there is a sequence $[f_n]$ of simple functions such that $f_n \xrightarrow{a.s.} f$ and $|f_n| \leq g \in \mathcal{L}^+$, so the conclusion follows by the DCT. ■

Multiplication Theorem

Let X be a separable Hilbert space, and suppose $X, Y : \Omega \rightarrow X$ are independent r.v.'s whose expectations exist. Then $EX \cdot Y$ exists, and

$$EX \cdot Y = EX \cdot EY.$$

Proof. The theorem is true for elementary functions, and if $X = \mathbb{R}$ and $X, Y \in \mathcal{M}^+$ the result follows by the MCT. But in any case $|X|$ and

$|Y|$ are independent by the Borel Functions Theorem, hence

$E|X||Y| = E|X|E|Y| < \infty$, so that $EX \circ Y$ exists. Now choose sequences X_n, Y_n of elementary functions such that X_n and Y_n are independent for each n , $|X_n| \leq |X| + 1$, $|Y_n| \leq |Y| + 1$, $X_n \xrightarrow{a.s.} X$, and $Y_n \xrightarrow{a.s.} Y$; and apply the DCT.

Note that if X is the complex plane and X, Y are independent, then X, Y^* are independent, $X \circ Y^* = XY$, and the theorem says that $EXY = EXEY$. This extends by induction to n factors.

Let \mathcal{G} be a σ -field such that $\mathcal{G} \subset \mathcal{A}$, and let $P_{\mathcal{G}}$ denote the restriction of P to \mathcal{G} . If $X : \Omega \rightarrow X$ is a Banach-valued, Bochner summable r.v., then the conditional expectation (c. expt.) of X given \mathcal{G} is an $(\mathcal{G}, \mathcal{B})$ measurable r.v., denoted by $E^{\mathcal{G}}X$ and determined up to a $P_{\mathcal{G}}$ equivalence by

$$\int_S (E^{\mathcal{G}}X) dP = \int_S X dP \quad (S \in \mathcal{G}).$$

This definition will be justified below.

In case $X = Z$, the existence and uniqueness (up to a $P_{\mathcal{G}}$ equivalence) of $E^{\mathcal{G}}X$ follows from the Radon-Nikodym Theorem applied to the P -continuous scalar measure $\phi : \mathcal{G} \rightarrow Z$ defined by $\phi S = \int_S X dP$. In fact, if $X = R$ and $EX = \infty$ exists, then ϕ is still P -continuous (but not necessarily σ -finite); in this case $E^{\mathcal{G}}X$ still exists by the extended R-N theorem (LOEVE, p. 133), but is not necessarily finite up to a $P_{\mathcal{G}}$ equivalence.

All equations, inequalities, and convergences involving conditional expectations are supposed to hold up to a $P_{\mathcal{G}}$ equivalence, but this will be understood rather than indicated each time. It will also

be assumed that all r.v.'s under consideration are $(\mathcal{A}, \mathcal{B})$ measurable, and that their expectations exist.

If $X = R$ and $0 \leq X_n \uparrow X$, then $0 \leq E^{\mathcal{B}} X_n \uparrow Y \in \mathcal{M}(\mathcal{B}, \mathcal{B})$, hence $\int_S X = \lim \int_S X_n = \lim \int_S E^{\mathcal{B}} X_n = \int_S Y$ by the usual MCT; therefore $Y = E^{\mathcal{B}} X$, so that the MCT holds (up to a $P_{\mathcal{B}}$ equivalence) for conditional expectations. The DCT in the real case follows from the MCT and the a.s. additivity of $E^{\mathcal{B}}$, and the complex case follows from the real case.

Denote by $\mathcal{L}^{\mathcal{B}}$ the family of all P -equivalence classes of $(\mathcal{A}, \mathcal{B})$ -measurable, summable r.v.'s X such that $E^{\mathcal{B}} X$ exists; and let $\mathcal{L}_{\mathcal{B}}$ be the family of all $P_{\mathcal{B}}$ -equivalence classes of $(\mathcal{B}, \mathcal{B})$ -measurable r.v.'s Y such that EY exists.

Conditional Expectations Theorem

(1) If $X \in \mathcal{L}^{\mathcal{B}}$ and $f: X \rightarrow Z$ is a continuous linear functional, then

$$E^{\mathcal{B}} fX = fE^{\mathcal{B}} X.$$

(2) If X is a separable Hilbert space, $X \in \mathcal{L}_{\mathcal{B}}$, $Y \in \mathcal{L}^{\mathcal{B}}$, and $\int |X| |Y| < \infty$, then

$$E^{\mathcal{B}} X \circ Y = X \circ E^{\mathcal{B}} Y.$$

(3) If $X \in \mathcal{L}^{\mathcal{B}}$, then $|E^{\mathcal{B}} X| \leq E^{\mathcal{B}} |X|$.

(4) $E^{\mathcal{B}}: \mathcal{L}^{\mathcal{B}} \rightarrow \mathcal{L}_{\mathcal{B}}$ is a continuous linear transformation.

(5) If $X_n \in \mathcal{L}^{\mathcal{B}}$, $|X_n| \leq g \in \mathcal{L}^+$, and $X_n \xrightarrow{\text{a.s.}} X$, then $X \in \mathcal{L}^{\mathcal{B}}$ and $E^{\mathcal{B}} X_n \rightarrow E^{\mathcal{B}} X$ (DCT); this implies that

$$\mathcal{L}^{\mathcal{B}} = \mathcal{L}.$$

Proof. Except for (2), the proof is patterned after the one in DRIML and HANS, who consider a real separable Banach space X .

(1) The proof of (1) follows immediately, since $fE^{\mathcal{B}}X$ is $(\mathcal{B}, \mathcal{B})$ -measurable and

$$\int_S E^{\mathcal{B}}X = \int_S X \text{ implies } \int_S fE^{\mathcal{B}}X = \int_S fX.$$

(2) First suppose $X = \mathbb{Z}$, and let $X = IA$, $A \in \mathcal{B}$. Then for each $S \in \mathcal{B}$,

$$\int_S (IA)E^{\mathcal{B}}Y = \int_{AS} E^{\mathcal{B}}Y = \int_{AS} Y = \int_S (IA)Y.$$

The assertion follows for simple functions X , and then for $X \in \mathcal{L}_{\mathcal{B}}$ by the DCT.

For let $[y_n]$ be a basis for X , and write

$$X_n = \sum_{k=1}^n (X \cdot y_k) y_k; \text{ so that } X_n \rightarrow X,$$

$$X_n \cdot Y = \sum_{k=1}^n (X \cdot y_k) (y_k \cdot Y) \rightarrow X \cdot Y,$$

and $|X_n \cdot Y| \leq |X| |Y|$. Since $X \cdot y_k$ is $(\mathcal{B}, \mathcal{B})$ measurable and $E^{\mathcal{B}}(y_k \cdot Y) = y_k \cdot E^{\mathcal{B}}Y$, it follows from the scalar case that

$$E^{\mathcal{B}}X_n \cdot Y = \sum_{k=1}^n (X \cdot y_k) E^{\mathcal{B}}(y_k \cdot Y) = X_n \cdot E^{\mathcal{B}}Y.$$

(3) First suppose $X = \mathbb{Z}$, and write $E^{\mathcal{B}}X = |E^{\mathcal{B}}X|e^{i\alpha}$, where $e^{i\alpha} = 1$ whenever $E^{\mathcal{B}}X = 0$. Then $e^{-i\alpha}$ is $(\mathcal{B}, \mathcal{B})$ -measurable, and using (2) it follows that

$$|E^{\mathcal{S}} S| = e^{-i\alpha} E^{\mathcal{S}} X = E^{\mathcal{S}} e^{-i\alpha} X = E^{\mathcal{S}} \operatorname{Re} [e^{-i\alpha} X] \leq E^{\mathcal{S}} |X| .$$

It will be shown in the next chapter, using the Hahn-Banach Theorem, that there exists a countable set $[f_n]$ of continuous linear functionals on X such that $|x| = \sup |f_n x|$ for every $x \in X$. In case X is a separable Hilbert space it suffices to take $f_n(x) = x \cdot u_n$, where $[v_n]$ is a countable dense set of nonzero vectors and $u_n = \frac{v_n}{|v_n|}$. Now using (1) and the scalar case,

$$|f_n E^{\mathcal{S}} X| = |E^{\mathcal{S}} f_n X| \leq E^{\mathcal{S}} |f_n X| \leq E^{\mathcal{S}} |X| ,$$

which yields

$$|E^{\mathcal{S}} X| = \sup |f_n E^{\mathcal{S}} X| \leq E^{\mathcal{S}} |X| .$$

(4) If $E^{\mathcal{S}} X$ and $E^{\mathcal{S}} Y$ exist, then it follows immediately from the definition that $\alpha E^{\mathcal{S}} X + \beta E^{\mathcal{S}} Y$ is a conditional expectation for $\alpha X + \beta Y$. To show that $E^{\mathcal{S}}: \mathcal{L}^{\mathcal{S}} \rightarrow \mathcal{L}_{\mathcal{S}}$ is a function, it must be shown that $E^{\mathcal{S}} X$ is uniquely determined (up to a $P_{\mathcal{S}}$ equivalence) whenever it exists. Suppose $Y \in \mathcal{L}_{\mathcal{S}}$ and $\int_S Y = \theta$ for all $S \in \mathcal{S}$, and let $[f_n]$ be the set of functionals in (3). Then using (1), $f_n \int_S Y = \int_S f_n Y = 0$, so that $f_n Y \stackrel{a.s.}{=} 0$ for each n by the scalar case, whence $|Y| = \sup |f_n Y| \stackrel{a.s.}{=} 0 \triangleright Y \stackrel{a.s.}{=} \theta$. This proves the uniqueness, so $E^{\mathcal{S}}$ is a linear transformation.

The spaces $\mathcal{L}^{\mathcal{S}}, \mathcal{L}_{\mathcal{S}} \subset \mathcal{L}$ are clearly NLS's with the usual norm $\|X\| = \int |X|$. Suppose $X_n \in \mathcal{L}^{\mathcal{S}}$ and $\|X_n\| \rightarrow 0$. Using (3),

$$\|E^{\mathcal{S}} X_n\| = \int |E^{\mathcal{S}} X_n| \leq \int E^{\mathcal{S}} |X_n| = \int |X_n| \rightarrow 0 .$$

(5) Let $f_n = \sup_{k \geq n} |X_k - X|$, so that $|f_n| \leq 2g$ and $f_n \downarrow 0$ a.s., hence $E^{\mathcal{B}} f_n \downarrow 0$ a.s. by the DCT for the scalar case. Thus if $m \leq n$, then

$$|E^{\mathcal{B}} X_n - E^{\mathcal{B}} X_m| \leq E^{\mathcal{B}} |X_n - X_m| \leq 2E^{\mathcal{B}} f_m \xrightarrow{\text{a.s.}} 0 \quad (m, n \rightarrow \infty),$$

so that $E^{\mathcal{B}} X_n \xrightarrow{\text{a.s.}} Y$ since X is complete. Since each $E^{\mathcal{B}} X_n$ is strongly $(\mathcal{A}, \mathcal{B})$ -measurable, it follows from the results of the previous chapter that Y is strongly $(\mathcal{A}, \mathcal{B})$ -measurable. It will be shown in the next chapter that every strongly $(\mathcal{A}, \mathcal{B})$ -measurable function Y is $(\mathcal{A}, \mathcal{B})$ -measurable in the sense that $Y^{-1}\mathcal{B} \subset \mathcal{A}$; in case X is a separable Hilbert space, this follows from

$$Y^{-1}[\alpha : |\alpha - \alpha_0| < \varepsilon] = [\omega : \sum |Y(\omega) \cdot y_n - \alpha_0 \cdot y_n|^2 < \varepsilon^2] \in \mathcal{A},$$

where $[y_n]$ is a basis.

Since $|E^{\mathcal{B}} X_n| \leq E^{\mathcal{B}} |X_n| \leq E^{\mathcal{B}} g$, it follows by the DCT for Bochner integrals that

$$\int_S Y = \lim \int_S E^{\mathcal{B}} X_n = \lim \int_S X_n = \int_S X,$$

therefore $E^{\mathcal{B}} X_n \xrightarrow{\text{a.s.}} E^{\mathcal{B}} X = Y$.

Finally, if $X \in \mathcal{L}$, then there exists a sequence of simple functions X_n such that

$$|X_n| \leq |X| + 1/\varepsilon \quad \text{and} \quad X_n \xrightarrow{\text{a.s.}} X.$$

But it follows easily that $E^{\mathcal{B}} X_n$ exists for each n , hence

$E^{\mathcal{B}} X \stackrel{\text{a.s.}}{=} \lim E^{\mathcal{B}} X_n$ exists by the DCT just established, and $\mathcal{L}^{\mathcal{B}} = \mathcal{L}$. ■

The following two classical lemmas will be needed to prove the Law of Large Numbers.

Toeplitz lemma. If

$$\sum_1^n a_k = b_n \uparrow \infty \quad \text{and} \quad y_n = \sum_1^n a_k x_k, \quad \text{then}$$

$$x_n \rightarrow x \triangleright \frac{y_n}{b_n} \rightarrow x$$

Proof. Suppose first that $x_n \rightarrow \theta$, let $\varepsilon > 0$, choose m so that $k \geq m \triangleright |x_k| < \varepsilon$, and let $\sum_1^m a_k |x_k| = M$. Then for $n > m$,

$$\left| \frac{y_n}{b_n} \right| \leq \frac{1}{b_n} \sum_{k=1}^m a_k |x_k| + \frac{1}{b_n} \sum_{k=m+1}^n a_k |x_k| \leq \frac{M}{b_n} + \varepsilon \rightarrow \varepsilon \quad (n \rightarrow \infty).$$

Thus $x_n \rightarrow \theta \triangleright \frac{y_n}{b_n} \rightarrow \theta$. Now if $x_n \rightarrow x$, then $x_n - x \rightarrow \theta$, whence

$$\frac{y_n}{b_n} - x = \frac{1}{b_n} \sum_1^n a_k (x_k - x) \rightarrow \theta.$$

Kronecker's lemma. If

$$b_n \uparrow \infty, \quad s_n = \sum_1^n x_k, \quad \text{and} \quad v \in X, \quad \text{then}$$

$$\sum \frac{x_n}{b_n} = v \triangleright \frac{s_n}{b_n} \rightarrow \theta.$$

Proof. Replace x_n by $b_n x_n$; and let $v_n = \sum_1^n x_k$, $v_0 = x_0 = \theta$,

and $s_n = \sum_1^n b_k x_k$. It suffices to show that $v_n \rightarrow v \triangleright \frac{s_n}{b_n} \rightarrow \theta$.

Summation by parts yields

$$s_n = \sum_1^n b_k (v_k - v_{k-1}) = b_n v_n - \sum_1^n (b_k - b_{k-1}) v_{k-1},$$

whence $\frac{s_n}{b_n} = v_n - \frac{y_n}{b_n} \rightarrow v - v = \theta$, applying the Toeplitz lemma with

$b_k - b_{k-1} = a_k$ and x_k replaced by v_{k-1} . ■

Throughout the remainder of this chapter, X is supposed to be a separable Hilbert space, and $\|X\|$ denotes that \mathcal{L}^2 norm of a r.v. X . Recall that \mathcal{L}^2 is also a Hilbert space with inner product defined by $X : Y = EX \cdot Y$.

Convergence of sequences and series in \mathcal{L}^2 norm will be denoted by $X_n \xrightarrow{2} X$ and $\sum X_n \stackrel{2}{=} X$ respectively. If X_n is an orthogonal sequence in \mathcal{L}^2 and $S_n = \sum_1^n X_k$, then

$$\|S_n - S_m\|^2 = \sum_{k=m+1}^n \|X_k\|^2,$$

hence $S_n - S_m \xrightarrow{2} \theta$ iff $\sum \|X_n\|^2 < \infty$. Furthermore, if $S_n \xrightarrow{2} X$, that is $X \stackrel{2}{=} \sum X_n$, then $\|X\|^2 = \sum \|X_n\|^2$.

Suppose $b_n \uparrow \infty$ and $\sum \frac{\|X_n\|^2}{b_n^2} < \infty$, so that $\sum \frac{X_n}{b_n} \stackrel{2}{=} Y$. Then $\frac{S_n}{b_n} \xrightarrow{2} \theta$ by Kronecker's lemma, hence

$$P \left[\left| \frac{S_n}{b_n} \right| \geq \epsilon \right] \leq \frac{\|S_n\|^2}{b_n^2 \epsilon^2} \rightarrow 0 \triangleright \frac{S_n}{b_n} \xrightarrow{P} \theta$$

by Tchebychev's inequality. Thus the following theorem holds.

Bernouilli-Tchebychev (Weak) Law of Large Numbers

If X_n are orthogonal r.v.'s and $b_n \uparrow \infty$, then

$$\sum \frac{\|x_n\|^2}{b_n^2} < \infty \Rightarrow \frac{s_n}{b_n} \xrightarrow{2} \theta \Rightarrow \frac{s_n}{b_n} \xrightarrow{P} \theta.$$

In particular,

$$\frac{1}{b_n^2} \sum_1^n \|x_k\|^2 \rightarrow 0 \Rightarrow \frac{s_n}{b_n} \xrightarrow{P} \theta.$$

If X is a r.v. whose expectation exists, then the variance of X is defined by

$$\sigma^2 X = \|X - EX\|^2 = \|X\|^2 - |EX|^2$$

A r.v. X is said to be centered at expectation iff $EX = \theta$; in this case $\sigma^2 X = \|X\|^2$.

Let X_n be summable Hilbert r.v.'s, $\mathfrak{F}_n = \left[\bigcup_{k=1}^{n-1} X_k^{-1} \mathfrak{B} \right]^-$ ($n=2,3,\dots$), $\mathfrak{F}_1 = [\emptyset, \Omega]$, $\xi_n = E^{\mathfrak{F}_n} X_n$ ($n=2,3,\dots$), and $\xi_1 = EX_1$; ξ_n are called the conditional expectations of X_n (given the predecessors), and X_n are said to be centered at conditional expectations iff $\xi_n \stackrel{a.s.}{=} \theta$. (This method is due to the famous French mathematician P. Levy.)

Suppose X_n are centered at c. expt.'s and $m < n$. Then X_n are centered at expt.'s ($EX_n = E\xi_n = \theta$) and $E^{\mathfrak{F}_n} X_m \cdot X_n = X_m \cdot E^{\mathfrak{F}_n} X_n = 0$ since X_m is \mathfrak{F}_n measurable; therefore $EX_m \cdot X_n = X_m : X_n = 0$.

Bienayme Equation

If X_n are centered at c. expt.'s, then they are centered at expt.'s and orthogonal, hence

$$\|s_n\|^2 = \sum_1^n \|x_k\|^2.$$

It should be noted that if X_n are independent, then $X_n^{-1} \mathcal{B}$ and \mathcal{G}_n are independent for each n ; and in this case $E^{\mathcal{G}_n} X_n \stackrel{a.s.}{=} EX_n$, so that centering at conditional expectations reduces to centering at expectations. For if $[y_k]$ is a basis for X , then $X_n \cdot y_k$ and IS are independent for each k and each $S \in \mathcal{G}_n$, using the Borel Functions Theorem; therefore $E(X_n \cdot y_k)IS = E(X_n \cdot y_k)PS$ by the Multiplication Theorem, which implies that $\int_S EX_n = (EX_n)PS = \int_S X_n$.

Since ξ_n is $(\mathcal{G}_n, \mathcal{B})$ measurable, so is $|\xi_n|^2$, and

$$\begin{aligned} E^{\mathcal{G}_n} |X_n - \xi_n|^2 &= E^{\mathcal{G}_n} |X_n|^2 - 2\operatorname{Re} E^{\mathcal{G}_n} \xi_n \cdot X_n + E^{\mathcal{G}_n} |\xi_n|^2 \\ &= E^{\mathcal{G}_n} |X_n|^2 - 2\operatorname{Re} \xi_n \cdot E^{\mathcal{G}_n} X_n + |\xi_n|^2 \\ &= E^{\mathcal{G}_n} |X_n|^2 - |\xi_n|^2, \end{aligned}$$

hence

$$\|X_n - \xi_n\|^2 = \|X_n\|^2 - \|\xi_n\|^2.$$

Now $|EX_n|^2 = |E\xi_n|^2 \leq (E|\xi_n|)^2 \leq \|\xi_n\|^2$, hence

$$\begin{aligned} \|X_n - \xi_n\|^2 &= \|X_n\|^2 - \|\xi_n\|^2 \leq \|X_n\|^2 - |EX_n|^2 \\ &= \|X_n - EX_n\|^2 = \sigma^2 X_n. \end{aligned}$$

That is, centering at condition expectations may modify variances, but they can only become smaller.

Kolmogorov's Inequality

If X_n are centered at c. expt.'s and $S_n = \sum_1^n X_k$, then

$$P \bigcup_{k=m}^n [|S_k - S_m| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} \sum_{k=m}^n \|X_k\|^2.$$

Proof. It suffices to assume that $m = 0$, where $S_0 = X_0 = \theta$. Let

$$A_k = \bigcap_{j=1}^k [|S_j| < \varepsilon] \quad (k = 1, \dots, n), \quad A_0 = \Omega,$$

and

$$B_k = A_{k-1} - A_k = \bigcap_{j=1}^{k-1} [|S_j| < \varepsilon] \cap [|S_k| \geq \varepsilon].$$

Then $A_k \downarrow$, hence $\sum_{k=1}^n B_k = A_0 - A_n = A_n^c = \bigcup_{j=1}^n [|S_j| \geq \varepsilon]$.

Consider $f(S_1, \dots, S_k) = S_k IB_k$, and note that

$$f(x_1, \dots, x_k) = x_k \left[\prod_{j=1}^{k-1} I[|x_j| < \varepsilon] \right] I[|x_k| \geq \varepsilon]$$

is a Borel function and each $S_j (j = 1, \dots, k)$ is \mathcal{G}_{k+1} measurable, where

$\mathcal{G}_{k+1} = \left[\bigcup_{j=1}^k X_j^{-1} \mathcal{B} \right]^-$; thus $S_k IB_k$ is also \mathcal{G}_{k+1} measurable. Hence

$$E^{\mathcal{G}_{k+1}} (S_k IB_k) \cdot (S_n - S_k) = S_k IB_k \cdot E^{\mathcal{G}_{k+1}} (S_n - S_k) = 0$$

since $E^{\mathcal{G}_{k+1}} X_j = E^{\mathcal{G}_j} X_j = \theta$ for $j \geq k+1$, which implies

$$E(S_k IB_k) \cdot (S_n - S_k) = \int_{B_k} S_k \cdot (S_n - S_k) = 0.$$

Therefore

$$\int_{B_k} |S_n|^2 = \int_{B_k} |S_n - S_k|^2 + 2 \operatorname{Re} \int_{B_k} (S_n - S_k) \cdot S_k + \int_{B_k} |S_k|^2 \geq \int_{B_k} |S_k|^2 \geq \varepsilon^2 P_{B_k},$$

so that

$$\varepsilon^2 {}^2P A_n^c = \varepsilon^2 \sum_1^n P B_k \leq \sum_1^n \int_{B_k} |S_n|^2 = \|S_n\|^2 = \sum_1^n \|x_k\|^2. \quad \blacksquare$$

Borel-Kolmogorov-Levy (Strong) Law of Large Numbers

(1) If X_n are Hilbert r.v.'s centered at c. expt.'s then

$$\sum \frac{\|x_n\|^2}{b_n^2} < \infty \triangleright \sum \frac{x_n}{b_n} \xrightarrow{a.s.} Y \triangleright \frac{S_n}{b_n} \xrightarrow{a.s.} \theta$$

(2) If X_n are independent and identically distributed Hilbert r.v.'s with common pr. distribution $Q = P_X$, where $E|X| < \infty$, then

$$\frac{S_n}{n} \xrightarrow{a.s.} EX$$

Proof. To prove (1), let $S_n' = \sum_1^n \frac{x_k}{b_k}$, and note that

$$P \bigcup_{k=m}^n [|S_k' - S_m'| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} \sum_{k=m}^n \frac{\|x_k\|^2}{b_k^2}$$

for every $\varepsilon > 0$ by Kolmogorov's inequality, hence

$$P \bigcup_{k=m}^{\infty} [|S_k' - S_m'| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} \sum_{k=m}^{\infty} \frac{\|x_k\|^2}{b_k^2} \rightarrow 0 \quad (m \rightarrow \infty)$$

by the Continuity Theorem. It follows from the Convergences Theorem

that $S_n' - S_m' \xrightarrow{a.s.} \theta$, so that $S_n' \xrightarrow{a.s.} Y$ for some Y since X is complete. Therefore $\frac{S_n}{b_n} \xrightarrow{a.s.} \theta$ by Kronecker's lemma.

Since $\|X_n - \xi_n\|^2 \leq \sigma^2 X_n$ it follows that in any case

$$\sum \frac{\sigma^2 x_n}{b_n^2} < \infty \triangleright \frac{1}{b_n} \sum_1^n (x_k - \xi_k) \xrightarrow{a.s.} \theta .$$

The proof of (2) will not be given here, since it is almost exactly like the classical proof (see LOEVE, p. 239). It should be noted that (2) is also true for Banach r.v.'s. GRENANDER, p. 144 shows that the proof can be reduced to the case of elementary r.v.'s, and then to the case of finite dimensional r.v.'s, which is covered by the Hilbert valued case for $X = Z^n$ since all finite dimensional topological linear spaces are topologically isomorphic to Z^n .

CHAPTER V

FUNCTIONAL ANALYSIS AND TOPOLOGY

In this chapter are presented those parts of topology and functional analysis which seem to be necessary for an understanding of the recent literature on generalized random variables. Among the results presented are two of the most celebrated and powerful theorems of modern analysis: the Hahn-Banach Theorem and the Tychonov Product Theorem. Although most of the theorems in this chapter are well known, their proofs will nevertheless be given for the sake of completeness.

For more comprehensive treatments of the topics presented here see AKHIESER [1], BANACH, DUNFORD and SCHWARTZ, HILLE and PHILLIPS, KOLMOGOROV and FOMIN, LIUSTERNIK and SOBOLEV, LOOMIS, NAIMARK, RIESZ and SZ. NAGY, TAYLOR, and ZAAANEN [2] for the functional analysis; and KELLEY for the topology. Excellent accounts of both topics can be found in McSHANE and BOTTS, and in SIMMONS. LOEVE also has a brief but lucid section on topology.

Functional Analysis

Throughout the present chapter, X denotes either a NLS or a topological space, or both. The meaning will always be clear from the context.

Let X and Y be two NLS's with the same scalar field Z . A transformation $T : X \rightarrow Y$ is said to be bounded iff it maps bounded sets in X onto bounded sets in Y . If T is semihomogeneous

($T\alpha x = \alpha Tx$ for $\alpha \geq 0$), then T is bounded iff the image of the unit sphere is bounded.

Boundedness Theorem. A semihomogeneous transformation T is bounded iff either

- (1) $|Tx| \leq \beta |x|$ for some constant $\beta \geq 0$, or
- (2) T is continuous at the origin θ .

Proof. The sufficiency of condition (1) is obvious. To prove the necessity, suppose there is no such constant β . Then for every $n = 1, 2, \dots$ there is an x_n in X such that $|Tx_n| > n^2 |x_n|$. Hence $|Tv_n| > n$, where $v_n = \frac{x_n}{n|x_n|}$, which implies that T is not bounded.

The necessity of condition (2) is obvious, since $|Tx| \leq \beta |x|$ clearly implies that T is continuous at θ . To prove the sufficiency, suppose T is not bounded. Then as above there is a sequence of vectors v_n such that $v_n \rightarrow \theta$, while $|Tv_n| > n$, hence T is not continuous at θ . Thus (1) and (2) are both equivalent to the boundedness of T . It follows that a linear transformation T , or a sublinear functional f , is continuous iff it is bounded.

The norm of a s. homog. transformation T is defined by $|T| = \inf [\beta : |Tx| \leq \beta |x|]$. Thus T is bounded iff $|T| < \infty$. It follows that

$$|T| = \sup_{|x|=1} |Tx|, \text{ and } |Tx| \leq |T| |x|.$$

To show this, note that $B = [\beta : |Tx| \leq \beta |x|] = [\beta : |Tu| \leq \beta \text{ for } |u| = 1]$. Now for $\beta \in B$, $|Tu| \leq \beta$ whenever $|u| = 1$, which implies that $|T|^* = \sup_{|u|=1} |Tu| \leq \beta$ and hence $|T|^* \leq \inf B = |T|$. But clearly $|Tu| \leq |T|^*$ whenever $|u| = 1$, so it follows that $|T|^* \in B$, and thus

$$|T| = \inf B \leq |T|^*.$$

The space of all bounded linear functionals $f : X \rightarrow Z$ is called the conjugate space (or adjoint space) of X . It will be denoted by X' , and its elements will be denoted by x' , y' or f , g , with or without indices, according to convenience. The conjugate space X' can be considered as a NLS, where the functionals $x' + y'$, $\alpha x'$ are defined by $(x' + y')x = x'x + y'x$, $(\alpha x')x = \alpha(x'x)$ respectively and the norm of x' is $|x'| = \sup_{|x|=1} |x'x|$. Clearly $|\alpha x'| = |\alpha| \cdot |x'|$, $|x' + y'| \leq |x'| + |y'|$, and $|x'| = 0 \triangleright x' = \theta'$, where $\theta'x \equiv 0$.

Note that X' is a space of functions $x' : X \rightarrow Z$, and thus an element $x' \in X'$ has two meanings: (1) x' is a vector in the NLS X' , and (2) x' is a scalar-valued function on the NLS X . For this reason it is necessary to distinguish between the norm $|x'|$ of x' considered as a vector, and the absolute value function associated with x' considered as a function, whose value at x is $|x'x|$. Thus the latter will be denoted by $|x'|^*$, so that $|x'|^* : X \rightarrow Z$ and $|x'|^*(x) \equiv |x'x|$. It will not be necessary to talk about the absolute value function $|x'|^*$ very often anyway, but its values $|x'x|$ do occur frequently. No confusion will arise if it is noted that x' , appearing by itself, always means a vector (which happens to be a scalar valued function), whereas $x'x = x'(x)$, appearing in this form, always means a scalar.

Conjugate Space Theorem. X' is always a Banach space.

Proof. Let $[x'_n]$ be a Cauchy sequence in X' . Then $|x'_m - x'_n| = \sup_{|x|=1} |x'_m x - x'_n x| \rightarrow 0$, hence $[x'_n x]$ is a Cauchy sequence for every unit $|x|=1$

vector x ; and in fact for every vector x since $x'_n \theta = 0$ and $x'_n x = |x| x'_n \frac{x}{|x|}$ for $x \neq \theta$. Therefore, since Z is complete, for each fixed x the sequence $[x'_n x]$ converges to some scalar, which will be denoted by $x' x$. It suffices to show that the functional x' , defined by $x' x = \lim x'_n x$, is linear and bounded, since $|x'_n - x'| = \sup_{|x|=1} |x'_n x - x' x| \rightarrow 0$. Now

$$x'(\alpha x + \beta y) = \lim x'_n(\alpha x + \beta y) = \lim (\alpha x'_n x + \beta x'_n y) = \alpha x' x + \beta x' y,$$

hence x' is linear. Note that $x'_n x \rightarrow x' x$ uniformly on the unit sphere $|x| = 1$, and $|x' x| \leq |x' x - x'_m x| + |x'_m x|$. Thus if m is chosen so that $|x' x - x'_m x| \leq 1$ uniformly for $|x| = 1$, then

$$|x'| = \sup_{|x|=1} |x' x| \leq \sup_{|x|=1} |x' x - x'_m x| + \sup_{|x|=1} |x'_m x| \leq 1 + |x'_m| < \infty. \quad \blacksquare$$

The zero functional θ' , defined by $\theta' x = 0$, is clearly in X' , but it is not so obvious that there exist others. For later purposes it is desirable to know that if x_0 is any point in X , then there exists a functional $f \in X'$ such that $f(x_0) = |x_0|$ and $|f| = 1$. This result is a corollary of the following celebrated theorem.

Hahn-Banach Theorem. Suppose X is a NLS, M is a linear manifold in X , and f_0 is a bounded linear functional on M . Then there exists a bounded linear functional $f \in X'$ such that:

- (1) $f(x) = f_0(x)$ for all x in M
- (2) $|f| = |f_0|$.

The real case was proved independently by H. HAHN (1927) and S. BANACH

(1929); then the complex case was proved by H. F. BOHNENBLUST and A. SOBEZYK (1938), also independently, using the real case. The proof is based on the following ingenious lemma.

Lemma. Suppose X is a real NLS, M is a linear manifold in X , and p is a sublinear functional on X . Let f_0 be a linear functional on M such that $f_0(x) \leq p(x)$, z any vector in $X - M$, and $M_z = [x + \alpha z : x \in M, \alpha \in \mathbb{R}]$. Then there exists a linear functional $f : M_z \rightarrow \mathbb{R}$ such that:

- (1) $f(x) = f_0(x)$ for all x in M ,
- (2) $f(x) \leq p(x)$ for all x in M_z .

Proof.

(a) The representation of a vector $v = x + \alpha z \in M_z$ is unique. For if $x + \alpha z = \theta$, then $\alpha z = -x \in M$, hence $\alpha = 0$, for otherwise $z = -\alpha^{-1}x \in M$. Thus if $x_1 + \alpha_1 z = x_2 + \alpha_2 z$, then $(x_1 - x_2) + (\alpha_1 - \alpha_2)z = \theta$, which implies that $\alpha_1 - \alpha_2 = 0$ and $x_1 - x_2 = \theta$.

(b) If $t \equiv f(z)$ and $f(x + \alpha z) \equiv f_0(x) + \alpha t$ ($x \in M, \alpha \in \mathbb{R}$), then f is linear on M_z and $f(x) = f_0(x)$ on M .

(c) Thus it suffices to choose t so that $f_0(x) + \alpha t \leq p(x + \alpha z) *$ for $x \in M$ and $\alpha \in \mathbb{R}$ (i.e. $f \leq p$ on M_z). This is clearly satisfied if $\alpha = 0$. If $\alpha > 0$, the condition $*$ is equivalent to $f_0(\alpha^{-1}x) + t \leq p(\alpha^{-1}x + z)$, and for $\alpha < 0$ it is equivalent to $f_0(\alpha^{-1}x) + t \geq -p(-\alpha^{-1}x - z)$, using the fact that p is semihomogeneous. Thus it suffices to show that a fixed t can be chosen so that

$$-p(-x-z) - f_0(x) \leq t \leq -f_0(y) + p(y+z) \text{ for all } x, y \text{ in } M.$$

(d) It suffices to show that $-p(-x-z) - f_0(x) \leq -f_0(y) + p(y+z)$, which is equivalent to $f_0(y) - f_0(x) = f_0(y-x) \leq p(y+z) + p(-x-z)$. But this is true by the subadditivity of p , therefore

$$t_0 = \sup_{x \in M} [-p(-x-z) - f_0(x)] \leq \inf_{y \in M} [-f_0(y) + p(y+z)] = t_1,$$

and any $t \in [t_0, t_1]$ will do for $f(z)$. This completes the proof of the lemma.

Now let \mathcal{E} be the class of all linear extensions $f: M_f \rightarrow \mathbb{R}$ of f_0 on linear manifolds $M_f (M \subset M_f \subset X)$, such that $f(x) \leq p(x)$ on M_f . Then \mathcal{E} is partially ordered by inclusion \subset if the functionals in \mathcal{E} are considered as sets of ordered pairs. Let ζ be any chain (linearly ordered subclass) in \mathcal{E} . Then $g = \bigcup \{f : f \in \zeta\} \supset f$ for all f in ζ , and g is a functional $g: D \rightarrow \mathbb{R}$ with domain $D \subset X$, such that $g(x) = f_0(x)$ on M and $g(x) \leq p(x)$ on D . Furthermore D is a linear manifold, which means that $g \in \mathcal{E}$. Therefore each chain ζ in \mathcal{E} has an upper bound, and by Zorn's lemma \mathcal{E} has a maximal element $f_1: M_{f_1} \rightarrow \mathbb{R}$. Finally $M_{f_1} = X$, for otherwise f_1 could be extended further by the lemma, which would contradict the maximality of f_1 .

Therefore, if X is a real linear space, p is a sublinear functional on X , M is a linear manifold in X , and f_0 is a linear functional on M such that $f_0(x) \leq p(x)$, then there is a linear functional $f: X \rightarrow \mathbb{R}$ such that

$$(1) \quad f(x) = f_0(x) \quad \text{for all } x \text{ in } M,$$

$$(2) \quad f(x) \leq p(x) \quad \text{for all } x \text{ in } X.$$

Now if $p(x) = |f_0| \cdot |x|$, then p is clearly sublinear on X , and $f_0(x) \leq |f_0(x)| \leq |f_0| |x| = p(x)$ for all x in M . Hence there

exists a linear functional $f : X \rightarrow \mathbb{C}$ such that $f(x) = f_0(x)$ on M and $f(x) \geq |f_0| \cdot |x|$. Furthermore $f(-x) = -f(x) \leq |f_0| |-x| = |f_0| \cdot |x|$, hence $f(x) \geq -|f_0| |x|$. Therefore $|f(x)| \leq |f_0| |x|$ for all $x \in X$, which implies that $|f| \leq |f_0|$. But clearly $|f_0| \leq |f|$, hence $|f| = |f_0|$.

Suppose that X is a complex NLS, M is a manifold in X , and $f_0(x) = g_0(x) + ih_0(x)$ is a complex bounded linear functional on M . Then $f_0(ix) = g_0(ix) + ih_0(ix) = if_0(x) = ig_0(x) - h_0(x)$, hence $h_0(x) = -g_0(ix)$, and $f_0(x) = g_0(x) - ig_0(ix)$.

Now g_0 is a real linear functional on M , considering X and M as real linear spaces for the moment, and $|g_0(x)| \leq |f_0(x)|$ implies $|g_0| \leq |f_0|$. Therefore, by the real case, there is a bounded real linear functional g on X such that $g(x) = g_0(x)$ on M and $|g| = |g_0| \leq |f_0|$. Define $f(x) = g(x) - ig(ix)$. Then clearly $f(x) = f_0(x)$ for all x in M , f is additive, and f is real linear since g is. But $f(ix) = g(ix) - ig(-x) = if(x)$, hence f is complex linear.

To complete the proof it suffices to show that $|f| \leq |f_0|$. Write $f(x) = |f(x)|e^{i\alpha}$, where $\alpha = \arg f(x)$. Then

$$|f(x)| = e^{-i\alpha} f(x) = f(e^{-i\alpha} x) = g(e^{-i\alpha} x).$$

For all x such that $|x| = 1$, $|e^{-i\alpha} x| = 1$, and

$$|f(x)| = g(e^{-i\alpha} x) \leq |g| = |g_0| \leq |f_0|.$$

This completes the proof of the Hahn-Banach Theorem. ■

Corollary 1. If x_0 is any vector in X , then there exists a bounded linear functional $f \in X'$ such that $f(x_0) = |x_0|$ and $|f| = 1$.

Proof. Let $M = [\alpha x_0 : \alpha \in \mathbb{Z}]$. Then clearly M is a linear manifold. Define the functional f_0 on M by $f_0(\alpha x_0) = \alpha |x_0|$. Then f_0 is a bounded linear functional on M , and $|f_0(\alpha x_0)| = |\alpha x_0|$ implies that $|f_0| = 1$ unless $x_0 = \theta$. Thus the conclusion follows immediately from the Hahn-Banach theorem unless $x_0 = \theta$. If $x_0 = \theta$, then choose any nonzero vector x_1 and repeat the same argument with x_0 replaced by x_1 . Thus there is a functional $f \in X'$ such that $f(\theta) = f(\theta) = 0 = |\theta|$ and $|f| = 1$. ■

Corollary 2. If X is a separable NLS, then there is a countable set $\{x'_n\} \subset X'$ such that $|x'_n| = 1$ and $|x| = \sup |x'_n x|$ for all $x \in X$. (If $\mathbb{Z} = \mathbb{R}$, then $|x| = \sup x'_n x$.)

Proof. Let $D = \{x_n\}$ be a countable dense subset of X . By Corollary 1, for each n there is an $x'_n \in X'$ such that $|x'_n| = 1$ and $x'_n x_n = |x_n|$. Now let x be any point in X , and let $\epsilon > 0$. Choose x_m so that $|x - x_m| < \epsilon$. Then $|x| - |x_m| < \epsilon$, and

$$|x'_m x_m| - |x'_m x| \leq |x'_m (x_m - x)| \leq |x'_m| \cdot |x_m - x| < \epsilon,$$

so $|x'_m x| \geq |x'_m x_m| - \epsilon = |x_m| - \epsilon > |x| - 2\epsilon$, which implies that $\sup |x'_n x| \geq |x|$. But clearly $\sup |x'_n x| \leq |x|$, since

$$|x'_n x| \leq |x'_n| \cdot |x| = |x|$$

for all n . ■

It follows easily from Corollary 2 that if X is a separable Banach space and $X : \Omega \rightarrow X$ is Bochner summable, then $\int_A X = 0$ for all $A \in \mathcal{A}$ implies $X \stackrel{a.s.}{=} \theta$. First note that if $x' \in X'$, then

$x' \int X = \int x' X$ for simple functions, hence also for summable functions since x' is continuous. If $[x'_n]$ is the set of functionals in Corollary 2, then $\int_A X = 0$ for all $A \in \mathcal{A}$ implies $x'_n \int_A X = \int_A x'_n X = 0$. Hence $x'_n X \stackrel{a.s.}{=} 0$ for each n by the scalar case, and then

$$|X| = \sup |x'_n X| \stackrel{a.s.}{=} 0$$

by Corollary 2, which implies that $X \stackrel{a.s.}{=} 0$.

Corollary 3. Suppose M is a linear manifold in X , $x \in M^{-C} = M^{CO}$, and $\delta = d(x, M) = \inf |x - y|$. Then there exists $x' \in X'$ such that $x'M = [0]$, $x'x = \delta$, and $|x'| = 1$. For the proof see HILLE and PHILLIPS p. 30.

The conjugate space of X' is denoted by X'' , and is called the second conjugate space of X . Thus X'' is the set of all bounded linear functionals $x'' : X' \rightarrow \mathbb{Z}$; it will be important in the considerations of this paper, inasmuch as it is used to define the weak topology on X to be denoted by \mathcal{W} . It is shown below that the original space X is always isometrically isomorphic to a subspace of X'' .

For each pair $x \in X$, $x' \in X'$, let $f(x, x') = F_x(x') = x'x$. If x' is held fixed and x is allowed to vary, this expression defines the functional $x' : X \rightarrow \mathbb{Z}$. On the other hand, x may be held fixed and x' allowed to vary; then $F_x(x') = x'x$ is a scalar for each $x' \in X'$, and thus defines a functional $F_x : X' \rightarrow \mathbb{Z}$. This functional F_x is actually an element of X'' , since

$$F_x(\alpha x' + \beta y') = (\alpha x' + \beta y')x = \alpha(x'x) + \beta(y'x) = \alpha F_x(x) + \beta F_x(y'),$$

and $|F_x(x')| = |x'x| \leq |x'| |x| = |x| |x'|$; clearly $|F_x| \leq |x|$.

Thus to each point x in X there is associated a point F_x in X'' by the mapping $F : X \rightarrow X''$, called the canonical mapping; where for each x in X , its image $F_x : X' \rightarrow E'$ under the canonical mapping is defined by

$$F_x(x') = x'x \quad \text{for } x' \in X'.$$

Clearly the canonical mapping is a homomorphism, $(F_{\alpha x + \beta y} = \alpha F_x + \beta F_y)$, since $F_{\alpha x + \beta y}(x') = x'(\alpha x + \beta y) = \alpha x'x + \beta x'y = \alpha F_x(x') + \beta F_y(x') = (\alpha F_x + \beta F_y)x'$. Thus it remains to show that F is 1-1 and isometric.

Since F is linear, to show that F is 1-1 ($F_x = F_y \triangleright x = y$) it suffices to show that $F_x = \theta'' \triangleright x = \theta$, where θ and θ'' are the origins (zero vectors) in X and X'' respectively. Suppose $x_0 \in X$ and $F_{x_0} = \theta''$. This means that $F_{x_0} \cdot x' = x'x_0 = 0$ for all $x' \in X'$. For each $x \in X$, by the first corollary of the Hahn-Banach theorem there is a functional $x' \in X'$ such that $|x'| = 1$ and $x'x = |x|$. Therefore $F_{x_0} = \theta'' \triangleright x'_0 x_0 = |x_0| = 0 \triangleright x_0 = \theta$. Notice also that given $x \in X$, there exists $x' \in X'$ such that $|x'| = 1$ and $|x| = x'x = |F_x x'| \leq |F_x|$. Therefore $|x| = |F_x|$ for all $x \in X$, since it was shown above that $|F_x| \leq |x|$. A NLS X is reflexive if and only if $FX = X''$. Clearly X must be a Banach space if it is reflexive, since X'' is a Banach space.

Extended Hahn-Banach Theorem. Let X be a NLS, M a linear manifold in X , Y a reflexive, separable Banach space, and $F_0 : M \rightarrow Y$ a bounded linear transformation. Then there exists a bounded linear transformation $F : X \rightarrow Y$ such that

$$(1) \quad Fx = F_0 x \quad \text{for all } x \text{ in } M$$

$$(2) \quad |F| = |F_0|.$$

Proof. For each $y' \in Y'$, $y' F_0 \in M'$, since

$$y' F_0(\alpha x) = \alpha y' (F_0 x), \quad y' F_0(x_1 + x_2) = y' (F_0 x_1 + F_0 x_2) = y' F_0 x_1 + y' F_0 x_2,$$

$$\text{and } |y' F_0(x)| \leq |y'| \cdot |F_0 x| \leq |y'| \cdot |F_0| \cdot |x|.$$

Therefore, by the Hahn-Banach Theorem, $y' F_0$ can be extended to a bounded linear functional $x' \in X'$ such that $|x'| = |y' F_0|$, where $x' = x'(y')$. (Note that $x'(\theta') = \theta'$.) Now define $F : X \rightarrow Y$ by the requirement that, for all $y' \in Y'$, $y' F(x) = x' x$, where $x' = x'(y')$.

Note that the functionals $x' = x'(y')$ are considered as fixed, once they have been chosen. Having fixed them, the equation $y' F(x) = x' x$ (where $x' = x'(y')$) determines $F(x) = y$ uniquely if it exists. For if $y' y_1 = y' y_2 = x' x$ for all y' in Y' , then $y_1 = y_2$. (As y' varies, so does x' , but x is fixed here.)

The question arises, however, as to whether y even exists so that $y' y = x' x$ for all y' in Y' , where x is fixed and $x' = x'(y')$. That is, given x in X and y' in Y' (which determines $x' = x'(y')$ in X'), y in Y must be chosen so that $y' y = x' x$. If $y' = \theta'$, then $x' = \theta'$, and there is nothing to prove. Otherwise, by the Hahn-Banach Theorem (Corollary 1) there exists y'' in Y'' such that $y'' y' = 1$. Then $(x' x) y'' \in Y''$ and $(x' x) y'' y' = x' x$. But since Y is reflexive, there exists $y \in Y$ such that $[(x' x) y''] y' = y' y = x' x$. Thus $y = F(x)$ is determined by

$$y' F(x) = x' x, \quad y' \in Y'.$$

Suppose $x \in M$. By the definition of x' , $y' F_0(x) = x' x$ for x in M , hence

$$F(x) = F_0(x) \text{ for all } x \text{ in } M.$$

$$\text{Since } y' F(x_1 + x_2) = x' (x_1 + x_2) = x' x_1 + x' x_2 = y' F(x_1) + y' F(x_2)$$

$$= y' (F x_1 + F x_2) \text{ for all } y' \in Y',$$

$$F(x_1 + x_2) = F(x_1) + F(x_2); \text{ similarly } F(\alpha x) = \alpha F(x).$$

It remains to show that $|F| \leq |F_0|$. It suffices to show that $|F x| \leq |F_0|$ for $|x| = 1$. Recall that for $|x| = 1$,

$$|x' x| = |y' F(x)| \leq |x'| = |y' F_0| = \sup_{|x|=1, x \in M} |y' F_0(x)|$$

for each y' in Y' .

Since Y is separable, there exists a countable set

$$[y'_n] \subset Y' \text{ such that } |y| = \sup |y'_n y|.$$

For $|x| = 1$ and each n ,

$$|y'_n F(x)| \leq \sup |y'_n F_0(x)| \leq \sup_{|x|=1} \sup_{x \in M} |y'_n F_0(x)| = \sup_{|x|=1, x \in M} |F_0 x| = |F_0|.$$

Therefore $|F(x)| = \sup_n |y'_n F(x)| \leq |F_0|$, which completes the proof. □

Topology. Let X be a topological space. A class \mathcal{B} is a subbase for the topology \mathcal{U} iff the class $\mathcal{I} = \left[\bigcap_{k=1}^n S_k : S_k \in \mathcal{B}; n = 1, 2, 3, \dots \right]$ of all finite intersections of \mathcal{B} -sets is a base for \mathcal{U} . Every class \mathcal{B} such that $\bigcup \mathcal{B} = X$ is a subbase for a topology \mathcal{U} on X , since every

$\mathcal{I}_x = [T : x \in T \in \mathcal{I}]$ is a nonempty class of nonempty sets which is closed under finite intersection, and hence every \mathcal{I}_x is a direction. This topology \mathcal{U} generated by \mathcal{I} is clearly the smallest topology on X containing \mathcal{I} , since $\mathcal{U} \subset \mathcal{V}$ for every topology \mathcal{V} on X such that $\mathcal{I} \subset \mathcal{V}$. That is $\mathcal{U} = \mathcal{I}$, where

$$\mathcal{I} = \bigcap [\mathcal{V} : \mathcal{V} \text{ is a topology, } \mathcal{I} \subset \mathcal{V}].$$

Conversely, if $\mathcal{U} = \mathcal{I}$, then \mathcal{I} is a subbase for \mathcal{U} . For otherwise the topology $\mathcal{U}(\mathcal{I})$ consisting of all unions of finite intersections of \mathcal{I} -sets is a topology on X containing \mathcal{I} such that $\mathcal{U} - \mathcal{U}(\mathcal{I}) \neq \emptyset$, which contradicts $\mathcal{U} = \mathcal{I} \subset \mathcal{U}(\mathcal{I})$.

Thus if $\bigcup \mathcal{I} = X$, then $\mathcal{U} = \mathcal{I}$ iff \mathcal{I} is a subbase for \mathcal{U} . It follows from the definition that $\mathcal{I} \subset \tilde{\mathcal{I}}$, $\tilde{\tilde{\mathcal{I}}} = \mathcal{I}$, $\mathcal{A} \subset \mathcal{I} \supset \mathcal{A} \subset \tilde{\mathcal{I}}$, $\tilde{\mathcal{A}} \subset \tilde{\mathcal{I}}$, and $(\mathcal{A} \cup \mathcal{I})^- = (\tilde{\mathcal{A}} \cup \tilde{\mathcal{I}})^-$. Thus the operation $(\tilde{})$ of taking minimal topologies containing classes \mathcal{I} with $\bigcup \mathcal{I} = X$ is analogous to the operation $(\bar{})$ of taking closures of sets A ; except that $(A \cup B)^- = \bar{A} \cup \bar{B}$, while in general $(\mathcal{A} \cup \mathcal{I})^- \neq \tilde{\mathcal{A}} \cup \tilde{\mathcal{I}}$, since unions of topologies are not in general topologies.

Let $(X_t, \mathcal{U}_t, \mathcal{B}_t)$ be a family of topological measurable (T.M.) spaces and $f_t : X_0 \rightarrow X_t$, $t \in T$, a family of functions, where X_0 is an arbitrary nonempty set. Recall that $\mathcal{U}_{0t} = f_t^{-1} \mathcal{U}_t$ is the smallest topology on X_0 such that f_t is $(\mathcal{U}_{0t}, \mathcal{U}_t)$ continuous, and $\mathcal{B}_{0t} = f_t^{-1} \mathcal{B}_t$ is the smallest σ -field on X_0 such that f_t is $(\mathcal{B}_{0t}, \mathcal{B}_t)$ measurable, for each $t \in T$. Let $\mathcal{I}_0 = \bigcup_{t \in T} f_t^{-1} \mathcal{U}_t$ and $\mathcal{B}_0 = \bigcup_{t \in T} f_t^{-1} \mathcal{B}_t$. Then clearly $\mathcal{U}_0 = \mathcal{I}_0$ is the smallest topology on X_0

such that f_t is $(\mathcal{U}_0, \mathcal{U}_t)$ continuous for all $t \in T$, and $\mathcal{B}_0 = \bar{\mathcal{B}}_0$ is the smallest σ -field on X_0 such that f_t is $(\mathcal{B}_0, \mathcal{B}_t)$ measurable for all $t \in T$. If \mathcal{U}_t and \mathcal{B}_t are replaced by \mathcal{J}_t and \mathcal{D}_t respectively in the definitions of \mathcal{J}_0 and \mathcal{D}_0 above, where $\mathcal{J}_t = \mathcal{U}_t$ and $\mathcal{D}_t = \mathcal{B}_t$, then \mathcal{U}_0 and \mathcal{B}_0 remain the same as before.

The topology $\mathcal{U}_0 = \mathcal{J}_0 = \left(\bigcup_{t \in T} \mathcal{U}_{0t} \right)^-$ is called the compound topology of the topologies \mathcal{U}_{0t} , and the σ -field $\mathcal{B}_0 = \bar{\mathcal{B}}_0 = \left(\bigcup_{t \in T} \mathcal{B}_{0t} \right)^-$ is called the compound σ -field of the σ -fields \mathcal{B}_{0t} .

The same situation can also be considered from a slightly different viewpoint. Namely consider the product space $X = \prod_{t \in T} X_t$, and recall that the projection mappings $p_t : X \rightarrow X_t$ are defined by $p_t x = x_t$ for each $x = (x_t : t \in T) \in X$. The family of functions f_t can be considered as a function

$$f = (f_t : t \in T) : X_0 \rightarrow X$$

with component functions f_t , and then

$$f_t = p_t f, \quad t \in T.$$

It would therefore be desirable to put a topology \mathcal{U} and a σ -field \mathcal{B} on X which are small enough so that $f : X_0 \rightarrow X$ is $(\mathcal{U}_0, \mathcal{U})$ continuous and $(\mathcal{B}_0, \mathcal{B})$ measurable, and large enough so that each projection $p_t : X \rightarrow X_t$ is $(\mathcal{U}, \mathcal{U}_t)$ continuous and $(\mathcal{B}, \mathcal{B}_t)$ measurable. This is always possible; in fact the projections p_t will be $(\mathcal{U}, \mathcal{U}_t)$ continuous and $(\mathcal{B}, \mathcal{B}_t)$ measurable if \mathcal{U} and \mathcal{B} are respectively the compound topology and compound σ -field on X generated by the projections, replacing X_0 by X and f_t by p_t in the

previous discussion.

Thus $\mathcal{U} = \mathcal{J}$, where $\mathcal{J} = \bigcup_{t \in T} p_t^{-1} \mathcal{U}_t$; and $\mathcal{B} = \mathcal{D}$, where $\mathcal{D} = \bigcup_{t \in T} p_t^{-1} \mathcal{U}_t$. But clearly $f^{-1}(p_t^{-1} U_t) = f_t^{-1} U_t \in \mathcal{U}_0$ for each $U_t \in \mathcal{U}_t$, and $f^{-1}(p_t^{-1} B_t) = f_t^{-1} B_t \in \mathcal{B}_0$ for each $B_t \in \mathcal{B}_t$, by definition of \mathcal{U}_0 and \mathcal{B}_0 . Hence $f^{-1} \mathcal{J} \subset \mathcal{U}_0$ and $f^{-1} \mathcal{D} \subset \mathcal{B}_0$, which implies that $f^{-1} \mathcal{J} \subset \mathcal{U}_0$ and $f^{-1} \mathcal{D} \subset \mathcal{B}_0$, as desired.

The topology \mathcal{U} is called the product topology on X , and the σ -field \mathcal{B} is called the product σ -field on X .

If $(\Omega, \mathcal{I}, \mathcal{A})$ is a given T.M. space and $X : \Omega \rightarrow X$, then it follows from the definitions that X is $(\mathcal{I}, \mathcal{U})$ -continuous or $(\mathcal{A}, \mathcal{B})$ -measurable iff X_t is $(\mathcal{I}, \mathcal{U}_t)$ -continuous or $(\mathcal{A}, \mathcal{B}_t)$ -measurable, respectively, for each $t \in T$.

Suppose (X, \mathcal{U}) is a top. space, \mathcal{J} is a base for \mathcal{U} , \mathcal{J} is a subbase for \mathcal{U} , $Y \subset X$, and $\mathcal{V} = \mathcal{U}Y = [UY : U \in \mathcal{U}]$. It follows that (Y, \mathcal{V}) is a topological space, called a subspace of (X, \mathcal{U}) ; and that $\mathcal{J}Y$ is a base for $\mathcal{U}Y$ and $\mathcal{J}Y$ is a subbase for $\mathcal{U} \circ Y$. Furthermore

$$\mathcal{V}^c = [Y - V : V \in \mathcal{V}] = \mathcal{U}^c Y,$$

and if $B \subset Y$ then $B_Y^- = B^- Y$, where B_Y^- is the closure of B with respect to \mathcal{V} . The topology \mathcal{V} is called the relative topology, or the relativization of \mathcal{U} on Y .

A top. space X is compact (K_2) iff every open cover of X has a finite subcover; note that every K_2 space is L_1 ($K_2 \subset L_1$). A subset $A \subset X$ is compact iff it is compact in its relative topology; equivalently iff every open cover of A has a finite subcover. It follows easily that every closed subset of a compact space is compact.

Alexander Subbase Theorem. Let (X, \mathcal{U}) be a top. space, and suppose \mathcal{B} is a subbase for \mathcal{U} . Then X is compact iff every open cover $\mathcal{Z} \subset \mathcal{B}$ has a finite subcover.

Proof. Let $A = [\mathcal{Z} \subset \mathcal{U} : \mathcal{Z} \text{ covers } X]$

$F = [\mathcal{Z} \subset \mathcal{U} : \text{some finite subclass of } \mathcal{Z} \text{ covers } X],$

$A' = [\mathcal{Z} \subset \mathcal{U} : \mathcal{Z} \text{ does not cover } X], \text{ and}$

$F' = [\mathcal{Z} \subset \mathcal{U} : \text{no finite subclass of } \mathcal{Z} \text{ covers } X].$

Then X is compact iff $A \subset F$, or equivalently $F' \subset A'$. Classes \mathcal{Z} in A , F , A' , and F' are called adequate, finitely adequate, inadequate, and finitely inadequate respectively. The plan of the proof is to show that if every finitely inadequate class of subbase sets is inadequate, then every finitely inadequate class is inadequate; symbolically,

$$F' \mathcal{B} \subset A' \mathcal{B} \supset F' \subset A'.$$

Consider a fixed finitely inadequate class $\mathcal{B} \in F'$, and let $F'_{\mathcal{B}} = [\mathcal{Z} \in F' : \mathcal{B} \subset \mathcal{Z}]$. To show that \mathcal{B} is inadequate, it suffices to show that $F'_{\mathcal{B}}$ has a maximal finitely inadequate class $\mathcal{M} \supset \mathcal{B}$, and that \mathcal{M} is inadequate. Let $F'_{\mathcal{B}}$ be partially ordered by inclusion \subset , and let \mathcal{C} be a linearly ordered subfamily (chain) of $F'_{\mathcal{B}}$. Then \mathcal{C} has an upper bound in $F'_{\mathcal{B}}$; namely $\mathcal{Z}^* = \bigcup [\mathcal{Z} : \mathcal{Z} \in \mathcal{C}]$. To see this, it suffices to show that $\mathcal{Z}^* \in F'_{\mathcal{B}}$. If this is false, there is a finite subclass $[C_1, \dots, C_n]$ of \mathcal{Z}^* which covers X , where $C_k \in \mathcal{Z}_k \in \mathcal{C}$. Then $\mathcal{Z} = \bigcup_1^n \mathcal{Z}_k \in F$, and $\mathcal{Z} \in \mathcal{C}$ since \mathcal{C} is linearly ordered, which is a contradiction. Therefore every chain in $F'_{\mathcal{B}}$ has an upper bound, and by

Zorn's lemma E'_\emptyset has a maximal class $\mathcal{M} = \mathcal{M}_\emptyset$. Note that \mathcal{M} is also a maximal class in E' , and $\emptyset \in \mathcal{M}$ since \mathcal{M} is maximal.

Since $\mathcal{M} \in E'$, clearly $\emptyset \cap \mathcal{M} = [S : S \in \emptyset, S \in \mathcal{M}] \in E'$, hence $\emptyset \cap \mathcal{M} \in \underline{A}'$ by hypothesis. Therefore, to show that $\mathcal{M} \in \underline{A}'$, it suffices to show that

$$\bigcup [M : M \in \mathcal{M}] \subset \bigcup [S : S \in \emptyset \cap \mathcal{M}] \neq X.$$

Let $x \in M \in \mathcal{M}$. Then by definition of subbase, there exist sets $S_j \in \emptyset$ ($j = 1, \dots, m$) such that

$$x \in \bigcap_1^m S_j \subset M.$$

Therefore it suffices to show that $S_j \in \mathcal{M}$ for some j , for then $x \in S_j \subset \bigcup [S : S \in \emptyset \cap \mathcal{M}]$. Suppose this is false, so that $S_j \notin \mathcal{M}$ for $j = 1, \dots, m$. Then $\mathcal{M}_j = \mathcal{M} + [S_j] \in E$ for each $j = 1, \dots, m$ by the maximality of \mathcal{M} ; thus for each j there is a finite subclass $[M_{jk} : k = 1, \dots, n_j] \subset \mathcal{M}$ such that

$$S_j \cup M_{j1} \cup \dots \cup M_{jn_j} = X.$$

Now for each j , $S_j = \bigcap_{i=1}^m S_i + \bigcup_{i=1}^m (S_j - S_i)$, and $S_j - S_i \subset S_i \cup M_{i1} \cup \dots \cup M_{in_i}$. It follows that

$$\bigcup_{j=1}^m S_j \cup M_{j1} \cup \dots \cup M_{jn_j} = \left(\bigcap_{i=1}^m S_i \right) \cup \left(\bigcup_{j=1}^m \bigcup_{k=1}^{n_j} M_{jk} \right) = X.$$

But since $\bigcap_{i=1}^m S_i \subset M \in \mathcal{M}$, this implies that $\mathcal{M} \in \underline{A}$, which is a contradiction. Thus the proof is complete. ■

Tychonov. Product Theorem. If (X_t, \mathcal{U}_t) , $t \in T$, is any family of compact top. spaces, then the product space (X^T, \mathcal{U}^T) is compact.

Proof. Let $\mathcal{B}^T = [p_t^{-1} U_t : U_t \in \mathcal{U}_t, t \in T]$ be the usual subbase for the product topology \mathcal{U}^T , and let ζ be any finitely inadequate class of subbase sets. According to the Alexander Subbase Theorem, it suffices to show that ζ is inadequate.

For each $t \in T$, let $\zeta_t = [U_t \in \mathcal{U}_t : p_t^{-1} U_t \in \zeta]$. Clearly each ζ_t is finitely inadequate for X_t ; hence also inadequate, since each X_t is compact. Therefore, for each $t \in T$, there exists a point x_t such that $x_t \notin U_t$ for all $U_t \in \zeta_t$. Thus if $x = (x_t : t \in T)$, then $x \notin p_t^{-1} U_t$ for all $p_t^{-1} U_t \in \zeta$, which means that ζ is inadequate. ■

If X is a NLS, then $X' \subset \prod_{x \in X} Z_x$, where $Z_x = Z$ for all x .

Thus it would seem that another natural topology on X' is the relativized product topology. It will be seen later that this topology, which is denoted by \mathcal{W}^* and called the weak* topology, does indeed turn out to be useful.

Recall that the usual subbase for the product topology \mathcal{U}^X on Z^X is

$$\mathcal{B}^X = \bigcup_{\substack{S \in \mathcal{J} \\ x \in X}} \{ [f \in Z^X : f_x \in S] \}$$

where \mathcal{J} is any subbase for the topology \mathcal{U} on Z . It will be convenient to take for \mathcal{J} the usual base for \mathcal{U} ; namely the class of all ε -neighborhoods. We could of course let \mathcal{J} be the usual countable base consisting of all ε -nhds. with rational centers and rational radii, but this is no help; the weak* topology is not in general second

countable, or even first countable, unless X is finite dimensional.

Thus the corresponding subbase for the weak topology \mathcal{W}^0 is

$$\mathcal{B}^0 = \bigcup_{\substack{\varepsilon > 0, \lambda \in \mathbb{Z} \\ x \in X}} \{ [x' \in X' : |x'x - \lambda| < \varepsilon] \},$$

and the sets

$$W_{x'_0}^0 = \bigcap_{k=1}^n [x' \in X' : |x'x_k - x'_0x_k| < \varepsilon]$$

form a base for the nhd. system of a point $x'_0 \in X'$.

\mathcal{W}^0 -Compactness Theorem. The unit ball $B_1^0 = [x' : |x'| \leq 1]$ is always \mathcal{W}^0 compact.

Proof. Since $|x'x| \leq |x'| \|x\| \leq |x|$ for all $x' \in B_1^0$,

$$B_1^0 = [x' : |x'| \leq 1] \subset \prod_{x \in X} [\alpha \in \mathbb{Z}_x : |\alpha| \leq |x|] \quad (\mathbb{Z}_x = \mathbb{Z} \text{ for all } x).$$

But $\prod_{x \in X} [\alpha \in \mathbb{Z}_x : |\alpha| \leq |x|]$ is compact in its relativized product topology (hence in the product topology on \mathbb{Z}^X) by the Tychonov Product Theorem, since $[\alpha \in \mathbb{Z} : |\alpha| \leq |x|]$ is a compact subset of \mathbb{Z} for each $x \in X$. Note here that the product topology on $\prod_{x \in X} [\alpha \in \mathbb{Z}_x : |\alpha| \leq |x|]$ is the same as the relativized product topology. Since a closed subset of a compact set is compact, it therefore suffices to show that B_1^0 is \mathcal{W}^0 closed.

Let $g \in \overline{B_1^0}$, the closure of B_1^0 in the product topology of \mathbb{Z}^X . It is not clear that g is even in X' , much less in B_1^0 ; this is what must be shown. Choose any $\varepsilon > 0$, and let x and y be any two points in X and α any scalar. Let

$$U_{g,\varepsilon} = [f : X \rightarrow Z; |f(x) - g(x)| < \varepsilon, |f(y) - g(y)| < \varepsilon, |f(x+y) - g(x+y)| < \varepsilon, \\ \text{and } |f(ax) - g(ax)| < \varepsilon].$$

Then $U_{g,\varepsilon}$ is a neighborhood of g ; therefore, since $g \in \overline{B_1^0}$, $U_{g,\varepsilon}$ contains some $f \in B_1^0$. Note that $f(x+y) = f(x) + f(y)$, $f(ax) = af(x)$, and $|f(x)| \leq |x|$. Hence,

$$|g(x) + g(y) - g(x+y)| \leq |g(x) - f(x)| + |g(y) - f(y)| + |f(x+y) - g(x+y)| < 3\varepsilon,$$

$$|ag(x) - g(ax)| \leq |ag(x) - af(x)| + |f(ax) - g(ax)| < (|a| + 1)\varepsilon, \text{ and}$$

$$|g(x)| \leq |g(x) - f(x)| + |f(x)| < \varepsilon + |x|.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $g(x+y) = g(x) + g(y)$, $g(ax) = ag(x)$, and $|g(x)| \leq |x|$ ($|g| \leq 1$), so that $g \in B_1^0$. Thus B_1^0 is closed, and the proof is complete. ■

The canonical mapping $F : X \rightarrow X''$, defined by $F_x(x') = x'x$, is an isometric isomorphism of X onto a subspace of X'' , namely the subspace consisting of all bounded linear functions $x'' : X' \rightarrow Z$ which are of the form $x''x' = x'x$ for some x in X . If X is only a NLS (not complete), then "subspace" should be replaced by "manifold."

The space X is said to be reflexive iff $FX = X''$. Examples of reflexive spaces are the scalar Lebesgue spaces L_μ^p ($1 < p < \infty$, μ σ -finite) and Hilbert space H .

It is sometimes convenient to identify X with its image FX under the canonical mapping, since in this way X can be considered as a subspace of the function space $X'' = (X')'$, which is in turn a subspace of the space $Z^{X'}$ of all scalar valued functions on X' . Thus,

$$X = FX \subset X'' = (X')' \subset Z^{X'} = \prod_{x' \in X'} Z_{x'}$$

One should realize here that X and FX are not really "equal" in the usual sense, but since they are isometrically isomorphic it will do no harm to identify them. In fact this identification makes it possible to introduce a new topology on X , namely the relativized product topology of $Z^{X'}$, called the weak topology on X and denoted by \mathcal{W} .

As before, the usual subbase for the product topology $\mathcal{U}^{X'}$ on $Z^{X'}$ is

$$\mathcal{S}^{X'} = \bigcup_{\substack{S \in \mathcal{I} \\ x' \in X'}} \{f \in Z^{X'} : f_{x'} \in S\}.$$

where \mathcal{I} is the class of all ε -neighborhoods in Z .

Thus the subbase for the relativized product topology on X'' is

$$\mathcal{S}'' = \bigcup_{\substack{\varepsilon > 0, \lambda \in Z \\ x' \in X'}} \{x'' \in X'' : |x''x' - \lambda| < \varepsilon\},$$

and the subbase for the relativized product topology \mathcal{W} on X is

$$\mathcal{S} = \bigcup_{\substack{\varepsilon > 0, \lambda \in Z \\ x' \in X'}} \{x \in X : |x'x - \lambda| < \varepsilon\}.$$

The base neighborhoods are finite intersections of subbase neighborhoods, and the class of all sets of the form

$$\mathcal{W}_{x_0} = \bigcap_{k=1}^n [x \in X : |x'_k x - x'_k x_0| < \varepsilon]$$

is a base for the nhd. system of the point x_0 .

It should be noted that, in constructing the weak topology \mathcal{W} , each element x in X can be thought of as a functional $x : X' \rightarrow Z$ (actually the functional F_x , where $F_x(x') = x'x$), and each element x' in X' can be thought of as a point. However, now that the weak topology has been constructed, one is free to reverse this and again think of elements x in X as points and elements x' in X' as functionals $x' : X \rightarrow Z$. With this interpretation it is clear, from the definition of base neighborhoods of a point x_0 above that the weak topology \mathcal{W} is the smallest topology on X with respect to which all bounded linear functionals $x' : X \rightarrow Z$ are continuous. In retrospect, it is equally clear from the definition that the \mathcal{W}' topology is the smallest topology on X' with respect to which all bounded linear functionals $x : X' \rightarrow Z$ (actually $F_x : X' \rightarrow E'$ where $F_x(x') = x'x$) are continuous.

Weak Compactness Theorem. If X is a reflexive Banach space, then the unit ball $B_1 = [x : |x| \leq 1]$ is weakly compact.

Proof. By the \mathcal{W}' Compactness Theorem the unit ball $B_1'' = [x'' : |x''| \leq 1]$ is compact with respect to the topology \mathcal{W}'' , where \mathcal{W}'' is the weak* topology on X'' . The topology \mathcal{W}'' is generated by the subbase

$$\mathcal{B}'' = \bigcup_{\substack{\varepsilon > 0, \lambda \in Z \\ x' \in X'}} \{[x'' : |x''x' - \lambda| < \varepsilon]\}.$$

But recall that the weak topology \mathcal{W} on X is generated by the subbase

$$\mathcal{J}'' = \bigcup_{\substack{\varepsilon > 0, \lambda \in \mathbb{Z} \\ x' \in X'}} \{ [x : |x'x - \lambda| < \varepsilon] \}$$

Therefore, since X is reflexive, the topological spaces (X, \mathcal{W}) and (X'', \mathcal{W}'') are clearly homeomorphic; the canonical mapping $F : X \rightarrow X''$ is a homeomorphism with respect to the weak topologies $\mathcal{W}, \mathcal{W}''$ as well as the strong topologies $\mathcal{U}, \mathcal{U}''$. Hence $B_1 = F^{-1}B_1''$ is compact. ■

It should be noted that the converse of this theorem is also true, so that a Banach space is reflexive if and only if the unit ball is weakly compact. For the proof see TAYLOR.

It was stated in the preceding two chapters that if X is a separable Banach space, then every weakly measurable function $X : \Omega \rightarrow X$ is $(\mathcal{A}, \bar{\mathcal{U}})$ measurable; this fact can now be proved with the aid of the Hahn Banach Theorem. If \mathcal{J} is the class of all ε -nhd.'s in Z and $\mathcal{J} = \bigcup_{f \in X'} f^{-1}\mathcal{J}$ is the usual subbase for the weak topology \mathcal{W} , then X is weakly measurable iff $X^{-1}\mathcal{J} = \bigcup_{f \in X'} (fX)^{-1}\mathcal{J} \subset \mathcal{A}$. Now since X is separable, $\bar{\mathcal{W}} = \bar{\mathcal{J}}$ using the Lindelöf Theorem; therefore X is weakly measurable iff it is $(\mathcal{A}, \bar{\mathcal{W}})$ measurable. It will be shown below that every strongly closed sphere is weakly closed. But every strongly open set is a countable union of strongly closed spheres, again using the Lindelöf Theorem, hence $\mathcal{U} \subset \bar{\mathcal{W}} \supset \bar{\mathcal{U}} = \bar{\mathcal{W}}$.

To show that every strongly closed sphere is weakly closed, it suffices to show that the unit ball B_1 is weakly closed, or equivalently $B_1^c = [x : |x| > 1]$ is weakly open. Suppose $|y| > 1$. By the

Hahn-Banach Theorem there exists $f \in X'$ such that $f(y) = |y|$ and $|f| = 1$, so that $|f(x)| \leq |x|$. Then $W = [x : |f(x) - |y|| < |y| - 1]$ is a weakly open nhd. of y and $W \subset B_1^c$.

It should be noted that the Hahn-Banach Theorem can also be used to show that the weak topology \mathcal{W} is Hausdorff. In the case of the conjugate space X' , it follows from the definition that (X', \mathcal{W}') is Hausdorff.

A topological space (X, \mathcal{U}) is said to be sequentially compact (K_1) iff every sequence in X has a convergent subsequence, and compact (K_2) iff every open cover of X has a finite subcover.

If X is C_1 and K_2 , then X is K_1 ($C_1 K_2 \subset K_1$). For suppose X is C_1 but not K_1 , and let x_n be a sequence in X with no convergent subsequence. For each $x \in X$, choose a nhd. V_x such that x_n is eventually in V_x^c ; this is possible since X is C_1 and x_n has no convergent subsequence. Then $\zeta = [V_x : x \in X]$ is an open cover of X which has no finite subcover, because x_n is eventually outside the union of any finite subclass of ζ by the way the sets V_x were chosen.

If X is C_2 and K_1 , then X is K_2 ($C_2 K_1 \subset K_2$). For suppose X is C_2 but not K_2 , and let ζ be an open cover with no finite subcover. By the Lindelöf Theorem there is a countable subcover, say $[U_n]$. Choose x_n inductively so that $x_n \in \bigcap_{k=1}^n U_k^c$; this is possible since $[U_n]$ has no finite subcover. Then x_n clearly has no convergent subsequence; because if $x \in U_m$, then $x_n \in U_m^c$ for all $n \geq m$, so that no subsequence can converge to x .

A subset $A \subset X$ is said to be compact or sequentially compact iff

it is a compact or sequentially compact subspace, respectively, in its relative topology. It is easy to see that a closed subset of a compact set is compact, with either notion of compactness.

Conversely a K_1 subset of a C_1T_2 space is closed, and a K_2 subset of a T_2 space is closed. Suppose X is C_1T_2 , and let $\emptyset \subsetneq S$ be K_1 . If $x \in S^-$, then there is a sequence $x_n \in S$ such that $x_n \rightarrow x$; and there is a subsequence x'_n such that $x'_n \rightarrow x' \in S$ since S is K_1 . But also $x'_n \rightarrow x$, hence $x = x' \in S$ since X is T_2 , which implies S is closed. Now suppose X is T_2 , and let $A \subsetneq X$ be K_2 and $y \in A^c$. For each $x \in A$ there exist open sets $U_x \in \mathcal{N}_x$ and $V_x \in \mathcal{N}_y$ such that $U_x V_x = \emptyset$; and $[U_x : x \in A]$ is an open cover of A , hence has a finite subcover $[U_{x_k} : k = 1, \dots, n]$. The sets $U = \bigcup_{k=1}^n U_{x_k} \in \mathcal{N}_A$, $V = \bigcap_{k=1}^n V_{x_k} \in \mathcal{N}_y$ are disjoint, thus A is closed.

A top. space X is locally compact (ℓK) iff each point has a closed compact nhd. Every locally compact Hausdorff space is regular. (See KELLEY, p. 141, Theorem 9 and p. 146, Theorem 17.)

Let X be a p -metric (M_0) space, $\epsilon > 0$, and $S \subsetneq X$. Then a set $E \subsetneq X$ is called an ϵ -net for S iff, for every $x \in S$, there exists a point $x_\epsilon \in E$ such that $d(x, x_\epsilon) < \epsilon$. The set S is totally bounded iff, for every $\epsilon > 0$, there exists a finite ϵ -net for S .

Every totally bounded p -metric space X is C_2 . To show this, for each $n = 1, 2, \dots$ choose a finite $\frac{1}{n}$ -net for X , say $[x_{n1}, \dots, x_{nk_n}] = E_n$; so that for each $x \in X$ there exists $x_{nj} \in E_n$ such that $x \in S(x_{nj}, \frac{1}{n})$. Then $\mathcal{B} = [S(x_{nj}, \frac{1}{n}) : j = 1, \dots, k_n; n = 1, 2, \dots]$ is a countable base; because for each $x \in X$ and $\epsilon > 0$, there is some

n such that $S(x, \frac{1}{n}) \subset S(x, \frac{\epsilon}{2})$, and then $x \in S(x_{n_j}, \frac{1}{n}) \subset S(x, \epsilon)$ for some $x_{n_j} \in E_n$.

A set S is conditionally compact iff every sequence in S has a convergent subsequence. Note that a set S is K_1 if it is closed and conditionally compact, and that S is totally bounded if and only if \bar{S} is totally bounded.

Hausdorff Compactness Theorem

If X is a complete p -metric space, then a set $S \subset X$ is conditionally compact iff it is totally bounded. A p -metric space is compact iff it is totally bounded and complete.

Proof. Suppose S is not totally bounded. Then for some $\epsilon > 0$, there is no finite ϵ -net for S . Choose any point $x_1 \in S$. Since $[x_1]$ is not an ϵ -net, $x_2 \in S$ can be chosen so that $d(x_1, x_2) > \epsilon$; since $[x_1, x_2]$ is not an ϵ -net, $x_3 \in S$ can be chosen so that $d(x_1, x_3), d(x_2, x_3) > \epsilon$. By the principle of inductive definition a sequence $[x_n] \subset S$ exists such that $d(x_m, x_n) > \epsilon$ for $m \neq n$, which implies that S is not conditionally compact. (Note that the necessity holds whether X is complete or not.)

Now suppose S is totally bounded, and let $[x_n]$ be any sequence in S . For each $k = 1, 2, \dots$ let $[a_{jk} : j = 1, \dots, p_k]$ be a $\frac{1}{k}$ -net for S . Then for each $k = 1, 2, \dots$

$$S \subset \bigcup_{j=1}^{p_k} S(a_{jk}, \frac{1}{k})$$

by definition of an ϵ -net. Note that it suffices to assume that $[x_n]$ has an infinite number of distinct points. Thus, for $k = 1$, at least

one of the spheres $S(a_{j1}, 1)$ say S_1 , contains an infinite subsequence $x_{1,n}$ of x_n ; for $k = 2$ some $S_2 = S(a_{j2}, \frac{1}{2})$ contains a subsequence $x_{2,n}$ of $x_{1,n}$, and so on. By the usual Cantor diagonal process, the diagonal sequence $x_{n,n}$ is a subsequence of $x_{k,n}$ for $k = 1, 2, \dots$. This implies that

$$x_{nn} \in S_n = S(a_{jk}, \frac{1}{k}) \quad \text{for } n \geq k,$$

hence $d(x_{nn}, x_{kk}) \leq d(x_{nn}, a_{jk}) + d(a_{jk}, x_{kk}) < \frac{2}{k}$. Therefore x_{nn} is a Cauchy sequence and, since X is complete, x_{nn} converges to at least one point in X , which implies that S is conditionally compact.

Evidently K_1 and K_2 compactness are equivalent for p -metric spaces, since every p -metric space is C_1 and every K_1 metric space is totally bounded, hence C_2 . To prove the second assertion, it suffices to show that every compact p -metric space is complete. But if X is compact and x_n is a Cauchy sequence, then there is a convergent subsequence $x_{n_k} \rightarrow x$; and $d(x_n, x) \leq d(x_n, x_{n_k}) + d(x_{n_k}, x) \rightarrow 0$ as $n \rightarrow \infty$. \square

CHAPTER VI

REGULAR MEASURES AND THE CONSISTENCY THEOREM

So far, the existence of the basic pr. space (Ω, \mathcal{A}, P) has been postulated, and the pr. distribution of a r.v. X in a measurable space $(\mathcal{X}, \mathcal{B})$ has been defined by $P_X B = P[X \in B]$. However, in pr. theory the domain space (Ω, \mathcal{A}, P) of a r.v. X is immaterial, except insofar as it provides a common frame of reference for all r.v.'s under consideration, and r.v.'s are used only as transformations from one pr. space to another, which preserve pr. distributions. Thus the only important thing about a r.v. X is its pr. distribution P_X . In fact, in practice the basic pr. space (Ω, \mathcal{A}, P) is unknown, hence even the r.v. X is unknown. This is because an actual physical experiment produces nothing but a pr. distribution Q on \mathcal{B} , or more generally a family $[Q_t : t \in T]$ of pr. distributions. If a single pr. distribution Q is involved, then there always exists a pr. space (Ω, \mathcal{A}, P) and a r.v. $X : \Omega \rightarrow \mathcal{X}$ such that $P_X = Q$; namely $(\Omega, \mathcal{A}, P) = (\mathcal{X}, \mathcal{B}, Q)$ and $X(x) = x$.

Consider a sequence $X = [X_n : n = 1, 2, \dots]$ of r.v.'s in a measurable space $(\mathcal{X}', \mathcal{B}')$, and write $X^n = [X_k : k = 1, \dots, n]$ and $(\mathcal{X}^n, \mathcal{B}^n) = \prod_{k=1}^n (\mathcal{X}_k, \mathcal{B}_k)$ for $n = 1, 2, \dots, \infty$, where $\mathcal{X}_k \equiv \mathcal{X}'$ and $\mathcal{B}_k \equiv \mathcal{B}'$. Define the projections $p_{nm} : \mathcal{X}^n \rightarrow \mathcal{X}_m$, $q_{nm} : \mathcal{X}^n \rightarrow \mathcal{X}^m$ ($m \leq n$) by $p_{nm}(x^n) = x_m$, $q_{nm}(x^n) = x^m$, and for $A^m \subset \mathcal{X}^m$ write $A^m \times \mathcal{X}^{n-m} = q_{nm}^{-1} A^m$; subscripts and superscripts $n = \infty$ will be dropped.

It follows easily that $q_{nm}q_n = q_m$, so that

$$A^m \times X^{-m} = q_m^{-1} A^m = q_n^{-1} q_{nm}^{-1} A^m = A^m \times X^{n-m} \times X^{-n}.$$

The sequence X can be regarded as a r.v. in the product space X , and its pr. distribution $P_0 = P_X$ on \mathcal{B} , defined by $P_0 B = P[X \in B]$, determines the joint pr. distributions $P_n = P_{X^n}$ on \mathcal{B}^n :

$$P^n B^n = P[X^n \in B^n] = P[X \in B^n \times X^{-n}] = P_0 B^n \times X^{-n}.$$

The pr. distributions P_n are consistent:

$$m < n, B^m \in \mathcal{B}^m \supset P^m B^m = P^n B^m \times X^{n-m} \equiv P^{nm} B^m.$$

Conversely, suppose consistent pr. distributions P^n on \mathcal{B}^n are given. The problem is to construct a pr. space (Ω, \mathcal{A}, P) and a sequence $X = [X_n : n = 1, 2, \dots]$ of r.v.'s $X_n : \Omega \rightarrow X'$ such that $P_{X^n} = P^n$. It suffices to construct a pr. Q on \mathcal{B} such that

$$Q^n B^n \equiv Q B^n \times X^{-n} = P^n B^n (B^n \in \mathcal{B}^n);$$

then take

$$(\Omega, \mathcal{A}, P) = (X, \mathcal{B}, Q) \text{ and } X_n(x) = x_n.$$

The classical Kolmogorov Consistency Theorem says this is always possible if X' is the real line and \mathcal{B}' is the Borel σ -field. But such a pr. Q does not always exist; for a counterexample see HALMOS [1], p. 214, prob. 3.

In this chapter, a generalization of the Kolmogorov Consistency Theorem will be proved. The proof leans heavily on some of the

topological concepts considered in the previous chapter, especially the concept of compactness. The proof also depends on the Caratheodory Extension Theorem and the concept of regular measures, which will be examined first.

So far in this paper, measures have been assumed to be given. But sometimes it is desirable to construct a measure $\bar{\mu}$ on a σ -field $\bar{\zeta}$ such that $\bar{\mu} = \mu$ on ζ , where μ is a given measure on ζ . If ζ is a ring such that X is a countable union of ζ sets, then the existence of such measures $\bar{\mu}$ is guaranteed by the following famous theorem due to Caratheodory.

Caratheodory Extension Theorem.

If μ is a σ -finite measure on a ring ζ , then there exists a unique measure $\bar{\mu}$ on $\bar{\zeta}$ such that $\bar{\mu} = \mu$ on ζ ; $\bar{\mu}$ is also σ -finite.

Proof. The existence proof is based on the concept of an outer measure.

A set function m^0 on a class \mathfrak{A} of subsets of X is an outer measure iff

- (1) $S_n \in \mathfrak{A}, \cup S_n \in \mathfrak{A} \triangleright m^0 \cup S_n \leq \sum m^0 S_n$
- (2) $A, B \in \mathfrak{A}; A \subset B \triangleright m^0 A \leq m^0 B$
- (3) $m^0 \emptyset = 0$

If m^0 is an outer measure on $\mathfrak{A}(X)$, then a set $A \subset X$ is said to be m^0 -measurable iff

$$m^0 S = m^0 (AS) + m^0 (A^c S) \text{ for every } S \subset X.$$

Lemma 1. If m^0 is any outer measure on $\mathfrak{A}(X)$ and \mathfrak{M}^0 is the class of all m^0 -measurable sets, then \mathfrak{M}^0 is a σ -field, and the

restriction of m^0 to \mathcal{M}^0 is a measure.

Lemma 2. If μ is a measure on a ring ζ such that X is the union of a countable number of ζ -sets, and if

$$\mu^0 S = \inf [\sum \mu A_n : A_n \in \zeta, \bigcup A_n \supset S]$$

for all $S \subset X$, then μ^0 is an outer measure on $\mathcal{S}(X)$. For the proofs of these lemmas, see LOEVE, p. 88.

Now let $m^0 = \mu^0$. Clearly $\mu^0 A = \mu A$ for $A \in \zeta$. Furthermore $\zeta \subset \mathcal{M}^0$. To see this, let $A \in \zeta$, $S \subset X$, and $\varepsilon > 0$, and choose $A_n \in \zeta$ so that $S \subset \bigcup A_n$ and

$$\mu^0 S + \varepsilon \geq \sum \mu A_n = \sum \mu A A_n + \sum \mu A^c A_n \geq \mu^0 AS + \mu^0 A^c S.$$

Since ε is arbitrary, $\mu^0 S \geq \mu^0 AS + \mu^0 A^c S$, and the opposite inequality follows by subadditivity. Thus $\bar{\zeta} \subset \mathcal{M}^0$, and it suffices to let $\bar{\mu}$ be the restriction of μ^0 to $\bar{\zeta}$.

To prove the uniqueness, let μ_1 and μ_2 be two extensions of μ to $\bar{\zeta}$ such that $\mu_1 = \mu_2 = \mu$ on ζ , choose $C_n \in \zeta$ so that $C_n \uparrow X$ and $\mu_1 C_n = \mu_2 C_n = \mu C_n < \infty$ for each n , and consider

$$\mathcal{M} = [A \in \bar{\zeta} : \mu_1 A C_n = \mu_2 A C_n \text{ for all } n].$$

Let $A_k \in \mathcal{M}$ be a monotone sequence, and $A = \lim A_k$. Then for each fixed n ,

$$\mu_1 A C_n = \lim_k \mu_1 A_k C_n = \lim_k \mu_2 A_k C_n = \mu_2 A C_n,$$

by the Continuity Theorem, so \mathcal{M} is a monotone class. Therefore $\zeta \subset \mathcal{M} \supset \bar{\zeta} = \zeta^S = \hat{\zeta} \subset \mathcal{M}$ by the Monotone Class Theorem, which means

$\bar{\zeta} = \mathcal{M}$. Now for any $A \in \bar{\zeta}$, $AC_n \uparrow A$, therefore

$$\mu_1 A = \lim \mu_1 AC_n = \lim \mu_2 AC_n = \mu_2 A$$

by the Continuity Theorem, which completes the proof. ■

Suppose μ is a content on a class ζ in a T.M. space $(X, \mathcal{U}, \mathcal{B})$, and let \mathcal{K} be the class of all closed compact subsets of X . Then a set $B \in \zeta$ is said to be

(1) outer regular iff $\mu B = \inf_{A \in \zeta, B \subset A^0} \mu A$

(2) inner regular iff $\mu B = \sup_{C \in \zeta, B \supset C^- \in \mathcal{K}} \mu C$

(3) regular iff both (1) and (2) hold.

The content μ is called a regular content iff every set $B \in \zeta$ is regular with respect to μ .

A semifield is a nonempty class \mathcal{D} of subsets of X such that, if $D, D' \in \mathcal{D}$, then

(1) $DD' \in \mathcal{D}$

(2) D^c can be written as a finite sum of \mathcal{D} -sets, say

$D^c = \sum_1^n E_j$, such that $D_k = D + \sum_1^k E_j \in \mathcal{D}$ for each partial sum ($k = 1, \dots, n$).

It follows that $X \in \mathcal{D}$, $\emptyset \in \mathcal{D}$, and

(2)' every proper difference $D' - D$ of two \mathcal{D} -sets ($D \subset D'$) can be written as a finite sum of \mathcal{D} -sets, say $D' - D = \sum_1^n E_j'$, such that $D_k' = D + \sum_1^k E_j' \in \mathcal{D}$ for each partial sum ($k = 1, \dots, n$).

A semiring is a nonempty class \mathcal{D} satisfying (1) and (2)', and it follows from (2)' that every semiring contains \emptyset , taking $D = D'$.

Clearly a semifield is a semiring containing \mathbb{X} .

Semiring Lemma

(1) Every additive set function μ on a semiring \mathfrak{D} is finitely additive.

(2) If μ is a measure (content) on a semiring \mathfrak{D} , then there exists a unique measure (content) μ' on the class \mathfrak{D}^r of all finite sums of \mathfrak{D} -sets such that $\mu' = \mu$ on \mathfrak{D} .

(3) A finite, regular content on a semiring \mathfrak{D} in a T.M. space is a measure on \mathfrak{D} .

Proof.

(1) It must be shown that if $D = \sum_1^m D_i$, where $D, D_i \in \mathfrak{D}$, then $\mu D = \sum_1^m \mu D_i$. This is true by hypothesis for $m = 2$; suppose it is true with m replaced by $m - 1$ ($m \geq 3$), and write $D = D_1 + \sum_1^n E_j$, where $E_j \in \mathfrak{D}$ and $D_1 + \sum_1^k E_j \in \mathfrak{D}$ for $k = 1, \dots, n$. It follows by a separate induction that $\mu D = \mu D_1 + \sum_1^n \mu E_j$. Furthermore

$$D_i = \sum_{j=1}^n D_i E_j \quad (i = 2, \dots, m),$$

where $D_i E_j \in \mathfrak{D}$ and $\sum_{j=1}^k D_i E_j \in \mathfrak{D}$ for each partial sum ($k = 1, \dots, n$),

hence also $\mu D_i = \sum_{j=1}^n \mu D_i E_j$; and $E_j = \sum_{i=2}^m D_i E_j$ ($j = 1, \dots, n$) implies

$\mu E_j = \sum_{i=2}^m \mu D_i E_j$ by the induction hypothesis. Therefore

$$\mu D = \mu D_1 + \sum_{j=1}^n \sum_{i=2}^m \mu D_i E_j = \mu D_1 + \sum_{i=2}^m \sum_{j=1}^n \mu D_i E_j = \mu D_1 + \sum_{i=2}^m \mu D_i .$$

(2) Suppose μ is a measure (content) on \mathfrak{D} , and for

$$D = \sum_1^m D_j = \sum_1^n E_k \in \mathfrak{D}^r \quad (D_j, E_k \in \mathfrak{D}) \text{ define}$$

$$\mu' D = \sum_1^m \mu D_j = \sum_1^m \sum_1^n \mu D_j E_k = \sum_1^n \mu E_k .$$

Note that μ' is well defined and uniquely determined, and $\mu' = \mu$ on \mathfrak{D} .

Let $D = \sum D_j$, where the sum is countable or finite according as μ is a measure or a content, and where $D = \sum_{i=1}^m E_i$,

$$D_j = \sum_{k=1}^{n_j} E_{jk} \quad (E_i, E_{jk} \in \mathfrak{D}).$$

Then

$$\mu' D = \sum_i \mu E_i = \sum_i \sum_j \sum_k \mu E_i E_{jk} = \sum_j \sum_k \sum_i \mu E_i E_{jk} = \sum_j \mu' D_j .$$

(3) Let $D = \sum D_n$, where $D, D_n \in \mathfrak{D}$. It was shown above that μ determines its extension to a content μ' on \mathfrak{D}^r ; for convenience in notation, write $\mu' = \mu$.

For each n , $\mu \sum_1^n D_j = \sum_1^n \mu D_j \leq \mu D$, hence $\sum \mu D_n \leq \mu D$.

Now let $\varepsilon > 0$ be given.

Since μ is inner regular, there is a set $C \in \mathfrak{D}$ such that $D \supset C^- \in \mathcal{K}$ and

$$\mu C > \mu D - \varepsilon .$$

Since μ is outer regular, for each n there is a set $E_n \in \mathfrak{D}$ such that $D_n \subset E_n^o$ and

$$\mu E_n < \mu D_n + \frac{\varepsilon}{2^n}.$$

Now $[E_n^o]$ is an open cover of C^- , hence there is a finite subcover, say $[E_k^o : k = 1, \dots, n]$. It follows that $C \subset \bigcup_{k=1}^n E_k$, whence

$$\mu D - \varepsilon < \mu C \leq \sum_1^n \mu E_k < \sum \mu D_n + \varepsilon,$$

which implies that $\mu D \leq \sum \mu D_n$ and μ is a measure. ■

It should be noted that the Extension Theorem is true for semirings instead of rings, using the Semiring Lemma (2).

If \mathfrak{E} is any class of subsets of X which contains X and \emptyset and is closed under finite unions and intersections, and \mathfrak{D} is the class of all proper differences of \mathfrak{E} -sets, then \mathfrak{D} is a semifield. For if $D = E - F$, $D' = E' - F' \in \mathfrak{D}$, then $DD' = EE' - EE'(F \cup F') \in \mathfrak{D}$; and $D^c = (EF^c)^c = F + E^c$ where $F, E^c \in \mathfrak{D}$ and $D + F = E \in \mathfrak{D}$. If \mathfrak{E} does not contain X , then \mathfrak{D} is still a semiring. For if $D \subset D'$, with the above notation, then $D' - D = [F - FF'] + [E' - (E \cup F')]$, and $D + [F - FF'] = E - FF' \in \mathfrak{D}$.

If \mathfrak{D} is a semifield and ζ is the class of all finite sums of \mathfrak{D} -sets, then ζ is a field. For if $A = \sum_1^m D_j \in \zeta$, $B = \sum_1^n E_k \in \zeta$, then clearly $AB = \sum_1^m \sum_1^n D_j E_k \in \zeta$; hence also $A^c = \bigcap_{j=1}^m D_j^c \in \zeta$, since each $D_j^c \in \zeta$ by axiom (2) and ζ is closed under pairwise (hence finite) intersection. If \mathfrak{D} is only a semiring, then ζ is still a ring.

Thus if \mathcal{E} is any class containing \emptyset, X which is closed under finite unions and intersections, and if \mathcal{D} and \mathcal{Z} are the classes defined above, then \mathcal{D} is a semifield and \mathcal{Z} is a field; furthermore

$$\mathcal{Z} = \mathcal{D} = \mathcal{E}.$$

(If $X \notin \mathcal{E}$, then a similar statement holds for semirings, rings, and minimal rings.) In particular, if (X, \mathcal{U}) is a top. space, \mathcal{D} is the class of all proper differences of closed sets, and \mathcal{Z} is the class of all finite sums of \mathcal{D} -sets, then

$$\mathcal{U}^c = \mathcal{Z} \supset \bar{\mathcal{U}} = \overline{\mathcal{U}^c} = \bar{\mathcal{Z}} = \hat{\mathcal{Z}}.$$

These relations will be useful in the proof of the Regular Measures Theorem. Recall that the same type of construction was used in Chapter IV, in the proof of the Independence Theorem.

A set A in a top. space X will be called a G_δ iff there exists a decreasing sequence U_n of open sets such that $A = \bigcap U_n$, and X will be called a G_δ -space iff every closed subset of X is a G_δ .

Every p -metric space is a G_δ -space: if A is closed and

$U_n = [x : d(x, A) = \inf_{y \in A} d(x, y) < \frac{1}{n}]$, then $A = \bigcap U_n$. Every regular L_2 space is a G_δ -space.

A set A in a T.M. space $(X, \mathcal{U}, \mathcal{B})$ will be called a $\mathcal{B}G_\delta$ iff $A \in \mathcal{B}$ and there is a sequence of open sets $U_n \in \mathcal{B}$ such that $U_n \downarrow A$.

A top. space X will be called σ -compact (σK) iff there exists an increasing sequence of closed compact sets C_n such that $X = \bigcup C_n$. For example, if X is any Banach space, then X' is σK with respect to the \mathcal{W}' topology, since all strongly closed spheres are \mathcal{W}' closed

and \mathcal{W}^0 compact. It follows easily that every σK space is L_1 , and every \mathcal{K} , L_1 space is σK . Furthermore, the product of a finite number of σK top. spaces is σK .

A T.M. space $(X, \mathcal{U}, \mathcal{B})$ will be called \bar{M}_{02} or σKG_δ iff (X, \mathcal{U}) is the same, respectively, and $\mathcal{B} = \bar{\mathcal{U}}$; and will be called $\Pi[\sigma KG_\delta]$ iff it is the product of a finite number of σKG_δ T.M. spaces.

Suppose $(X, \mathcal{U}, \mathcal{B})$ is a T.M. space, \mathcal{K} is the class of all closed compact subsets of X , and μ is a measure on \mathcal{B} . Then X is said to be μ -a.s. σ -compact iff there exists a sequence $C_n \in \mathcal{BK}$ such that $C_n \uparrow C$ and $\mu C^c = 0$. (In the terminology of HALMOS [1], p. 74, this means there exists an increasing sequence of closed compact measurable sets whose union is a thick set with respect to μ .) A T.M. space will be called a.s. σ -compact (a.s. σK) iff it is μ -a.s. σ -compact with respect to every finite measure μ on \mathcal{B} .

The following theorem is a problem in HALMOS [1], p. 40 (for metric spaces).

Theorem. Let X be a separable, complete p -metric space (\bar{M}_{02} space) with metric topology \mathcal{U} , and let μ be a finite measure on $\mathcal{B} = \bar{\mathcal{U}}$. Then there exists a sequence of closed compact sets C_n such that $C_n \uparrow C$ and $\mu C = \mu X$. Therefore every \bar{M}_{02} space is a.s. σK .

Proof. It suffices to assume that $\mu X = 1$.

Let $[x_n]$ be a countable dense set, and $0 < \varepsilon < 1$. For each $k, m = 1, 2, \dots$, define

$$F_{km} = \bigcup_{n=1}^m \bar{S}(x_n, \frac{1}{k})$$

Let m_1 be the smallest positive integer such that $\mu F_{1m_1} > 1 - \epsilon$. Clearly m_1 exists, because $F_{1m} \uparrow X$ since $[x_n]$ is dense, which implies that $\mu F_{1m} \uparrow \mu X = 1$ by the Continuity Theorem. Similarly, $F_{1m_1} F_{2m} \uparrow F_{1m_1}$ ($m \rightarrow \infty$), hence

$$\mu F_{1m_1} F_{2m} \uparrow \mu F_{1m_1} > 1 - \epsilon.$$

Let m_2 be the smallest positive integer such that

$$\mu F_{1m_1} F_{2m_2} > 1 - \epsilon.$$

Proceeding by induction, let m_k be the smallest positive integer such that

$$\mu \bigcap_{j=1}^k F_{jm_j} > 1 - \epsilon \quad (k = 1, 2, \dots).$$

Let $D_k = \bigcap_{j=1}^k F_{jm_j} \downarrow$, and $D = \bigcap D_k = \bigcap F_{km_k} = \bigcap_k \bigcup_{n=1}^{m_k} \bar{S}(x_n, \frac{1}{k})$, so that $\mu D = \lim \mu D_k \geq 1 - \epsilon$ by the Continuity Theorem. Furthermore, D is closed; and D is compact by the Hausdorff Theorem since $[x_n]_1^{m_k}$ is a $\frac{1}{k}$ -net for D for $k = 1, 2, \dots$. For if $x \in D$, then for each k , $x \in \bar{S}(x_n, \frac{1}{k})$ for some $n = 1, \dots, m_k$.

Now choose for each n a closed compact set D_n such that $\mu D_n \geq 1 - \frac{1}{n}$, and let $C_n = \bigcup_{k=1}^n D_k$ and $C = \bigcup C_n$. Then each C_n is closed and compact, and $\mu C_n \geq \mu D_n \geq 1 - \frac{1}{n}$, hence $\mu C = \lim \mu C_n = 1$ by the Continuity Theorem. ■

Suppose $(X, \mathcal{U}, \mathcal{B})$ is a T.M. space, \mathcal{K} is the class of all closed compact subsets of X , and μ is a measure on \mathcal{B} . A set $B \in \mathcal{B}$ is

(1) outer regular iff $\mu B = \inf_{B \subset A \in \mathcal{B}} \mu A$

(2) inner regular iff $\mu B = \sup_{B \supset C \in \mathcal{B}} \mu C$

(3) regular iff both (1) and (2) hold.

A class $\zeta \subset \mathcal{B}$ is regular iff every set $B \in \zeta$ is regular, and the measure μ is regular iff \mathcal{B} is regular.

It will be shown that if X is either \bar{M}_{02} or $\Pi[\sigma KG_\delta]$, then every finite measure μ on \mathcal{B} is regular. The proof is based on the methods of HALMOS [1], p. 224-228, who deals with locally compact Hausdorff spaces.

Lemma 1.

(a) The union of any sequence of outer regular sets is outer regular, and the union of an increasing sequence of inner regular sets is inner regular.

(b) If μ is finite, then the intersection of any sequence of inner regular sets is inner regular, and the intersection of a decreasing sequence of outer regular sets is outer regular.

Proof. (a) Let $[B_n]$ be any sequence of outer regular sets, $B = \bigcup B_n$, and $\varepsilon > 0$. If $\mu B = \infty$, then B is trivially outer regular. Assume $\mu B < \infty$, and for each n choose an open measurable set $V_n \supset B_n$ such that $\mu V_n < \mu B_n + \frac{\varepsilon}{2^n}$. Let $V = \bigcup V_n \supset B$. Then

$$V - B = \bigcup_m [V_m \cdot \bigcap_n B_n^c] \subset \bigcup_m V_m B_m^c = \bigcup V_m - B_m,$$

therefore $\mu V - \mu B \leq \sum \mu V_m - \mu B_m \leq \sum \frac{\varepsilon}{2^m} = \varepsilon \triangleright B$ is outer regular.

Now suppose $[B_n]$ is an increasing sequence of inner regular sets,

and let $B = \bigcup B_n$. Then $\mu B_n \uparrow \mu B$ by the Continuity Theorem. Let $r < \mu B$, choose m so that $r < \mu B_m \leq \mu B$, then choose a closed compact measurable set C so that $C \subset B_m \subset B$ and $\mu C > r$. Thus B is inner regular. The proof of (b) is similar (see HALMOS [1], p. 226). ■

Lemma 2. Every finite sum of regular sets is regular. If μ is finite and ζ is any class of closed (open) regular sets, then every proper difference of two ζ sets is regular.

Proof. It follows from lemma 1 that every finite sum of outer regular sets is outer regular; and it follows immediately from the definition that every finite sum of inner regular sets is inner regular, since a finite union of closed compact measurable sets also has these properties.

The second assertion will be proved for closed sets; the proof for open sets is similar. Let $A, B \in \zeta$, $A \supset B$, and $\varepsilon > 0$. Since A is inner regular, there is a closed compact measurable set $C \subset A$ such that $\mu C > \mu A - \varepsilon$; and since B is outer regular, there is an open measurable set $V \supset B$ such that $\mu V < \mu B + \varepsilon$. Then $C - V \subset A - B$, $C - V$ is a closed compact measurable set, and

$$\mu(C - V) = \mu C - \mu CV > \mu A - \varepsilon - \mu B - \varepsilon = \mu(A - B) - 2\varepsilon;$$

therefore $A - B$ is inner regular. Since A is outer regular, there is an open set U such that $\mu U < \mu A + \varepsilon$. Thus $U - B$ is open, $A - B \subset U - B$, and $\mu(U - B) < \mu(A - B) + \varepsilon$, which means $A - B$ is outer regular. ■

Regular Measures Theorem

Let $(X, \mathcal{U}, \mathcal{B})$ be an a.s. σK T.M. space; and suppose ζ is a class of closed $\mathcal{B}G_\delta$'s such that ζ is closed under finite intersections, $\emptyset, X \in \zeta$, and $\mathcal{B} = \bar{\zeta}$. Then every finite measure μ on \mathcal{B} is regular.

In particular, if X is either \bar{M}_{02} or $\Pi[\sigma KG_\delta]$, then every finite measure μ on \mathcal{B} is regular.

Proof. Let μ be a finite measure on \mathcal{B} , and choose $C_n \in \mathcal{BK}$ so that $C_n \uparrow C$ and $\mu C = \mu X$. If $B \in \mathcal{B}$ is closed, then $BC_n \in \mathcal{BK}$ for each n , and $BC_n \uparrow BC \triangleright \mu BC_n \uparrow \mu BC = \mu B$ by the Continuity Theorem; thus every closed measurable set is inner regular. Since μ is finite, it follows by the Continuity Theorem that every $\mathcal{B}G_\delta$ is outer regular.

If ζ^* is the class of all finite unions of ζ sets, then ζ^* is regular by the above remarks and lemma 1 (a). Furthermore, ζ^* is closed under finite unions and finite intersections, so that the class \mathcal{D} of all proper differences of ζ^* sets is a semifield; and \mathcal{D} is regular by lemma 2.

The class \mathcal{F} of all finite sums of \mathcal{D} sets is a field (in fact $\mathcal{F} = \bar{\zeta}$), and \mathcal{F} is regular by lemma 2. But the class \mathcal{R} of all regular sets is a monotone class by lemma 1, hence $\mathcal{F} \subset \mathcal{R}$ implies

$$\mathcal{B} = \bar{\mathcal{F}} = \hat{\mathcal{F}} \subset \mathcal{R}$$

by the Monotone Class Theorem.

To prove the particular assertion in case X is \bar{M}_{02} , simply take $\zeta = \mathcal{U}^c$, so that $\bar{\zeta} = \bar{\mathcal{U}} = \mathcal{B}$; and recall that every \bar{M}_{02} space

is a.s. σK , and every closed subset of a M_0 space is a G_δ .

Suppose $(X, \mathcal{U}, \mathcal{B}) = \prod_{i=1}^m (X_i, \mathcal{U}_i, \mathcal{B}_i)$ is the product of a finite number of $\sigma K G_\delta$ spaces, where $\mathcal{B}_i = \bar{\mathcal{U}}_i$, and let $\zeta = [\prod_{i=1}^m A_i : A_i \in \mathcal{U}_i^c]$. It follows easily that ζ is closed under finite intersections, $\mathcal{B} = \bar{\zeta}$, and each ζ set is a $\mathcal{B} G_\delta$. For each $i = 1, \dots, m$, let C_{in} be a sequence of closed compact sets such that $C_{in} \uparrow X_i$. Then $C_n = \prod_{i=1}^m C_{in} \in \mathcal{BK}$ and $C_n \uparrow X$, so that X is σK . ■

Let $(X_n, \mathcal{U}_n, \mathcal{B}_n)$ be a sequence of T.M. spaces, and consider the product spaces $(X^n, \mathcal{U}^n, \mathcal{B}^n) = \prod_{k=1}^n (X_k, \mathcal{U}_k, \mathcal{B}_k)$ for $n = 1, 2, \dots, \infty$; the notation introduced at the beginning of the chapter will be used. Let $\mathcal{K}_n, \mathcal{K}^n$ be the classes of all closed compact subsets of X_n, X^n respectively.

Kolmogorov Consistency Theorem

Consistent, inner regular pr.'s P^n on \mathcal{B}^n determine a unique pr. P on \mathcal{B} such that

$$PB^n \times X^{-n} = P^n B^n \quad (B^n \in \mathcal{B}^n),$$

provided that the top. spaces (X_n, \mathcal{U}_n) are regular.

The inner regularity hypothesis is satisfied in particular whenever the T.M. spaces $(X_n, \mathcal{U}_n, \mathcal{B}_n)$ are either \bar{M}_{02} or $\sigma K G_\delta$, where $\mathcal{B}_n = \bar{\mathcal{U}}_n$.

Proof. Let ζ be the class of all "Borel cylinders" of the form $B^n \times X^{-n}$, where $B^n \in \mathcal{B}^n$. It follows from the definitions that ζ is a field and $\mathcal{B} = \bar{\zeta}$. Define P on ζ by $PB^n \times X^{-n} = P^n B^n, B^n \in \mathcal{B}^n$; it follows from the consistency hypothesis that P is well defined and additive on ζ .

Therefore, by the Caratheodory Extension Theorem, it suffices to show that P is a measure on ζ . By the Continuity Theorem, it suffices to show that P is continuous from above at \emptyset . That is, if $A_n \in \zeta$ and $A_n \downarrow A \in \zeta$, then $A = \emptyset \Rightarrow PA_n \downarrow 0$, or equivalently

$$PA_n > 2\varepsilon > 0 \Rightarrow A = \bigcap A_n \neq \emptyset.$$

It suffices to assume that $A_n = B^n \times X^{-n}$, $B^n \in \mathcal{B}^n$, since in any case there exists a sequence of sets A'_n of this form such that $\bigcap A_n = \bigcap A'_n$.

By the inner regularity hypothesis, for each n there exists a closed compact set $D^n \in \mathcal{B}^n$ such that $D^n \subset B^n$ and $P^n(B^n - D^n) < \varepsilon 2^{-n}$. Thus if $C_n = D^n \times X^{-n}$, then $P(A_n - C_n) = P^n(B^n - D^n) < \varepsilon 2^{-n}$.

Consider $E_n = \bigcap_{k=1}^n C_k \subset C_n \subset A_n$, and note that

$$A_n - E_n = \bigcup_{k=1}^n A_n C_k^c \subset \bigcup_{k=1}^n (A_k - C_k),$$

so that

$$P(A_n - E_n) \leq \sum_{k=1}^n P(A_k - C_k) < \sum_{k=1}^{\infty} \varepsilon 2^{-k} = \varepsilon.$$

Thus $2\varepsilon < PA_n = P(A_n - E_n) + PE_n < \varepsilon + PE_n \Rightarrow PE_n > \varepsilon$.

It follows that $E_n \neq \emptyset$ for each fixed n ; let $x = (x^n, x^{-n})$ be a point in E_n (x depends on n here), and note that $x^{-n} = (x_{n+1}, x_{n+2}, \dots)$ can be chosen arbitrarily.

Since $x \in C_k = D^k \times X^{-k}$ for $k = 1, \dots, n$, clearly $x_k \in D_k \equiv p_{kk} D^k \subset X_k$; thus x can be chosen so that $x_k \in D_k$ for all k , since x_k is arbitrary for $k > n$.

Now the projection mapping $p_{kk} : X^k \rightarrow X_k$ is continuous and D^k is compact in X^k , hence D_k is compact in X_k .

But since each X_k is regular, D_k^- is also compact (see KELLEY, p. 141, Theorem 10 and p. 161, prob. B(b)), hence $D = \prod_{k=1}^{\infty} D_k^-$ is closed and compact, using the Tychonoff Product Theorem.

Thus DE_n is a decreasing sequence of nonempty, closed compact sets, which implies that $\bigcap DE_n \neq \emptyset$. But $DE_n \subset A_n$, hence $\bigcap A_n \neq \emptyset$.

It should be noted that if each X_k is C_1T_2 , then compact (K_2) can be replaced by K_1 . In this case, each D_k is K_1 , hence also closed. It follows that $D = \prod_{k=1}^{\infty} D_k$ is closed, and D is K_1 by the Cantor diagonal method. Thus DE_n is a decreasing sequence of nonempty, closed K_1 sets, and it is still true that $\bigcap DE_n \neq \emptyset$.

The particular assertion follows from the Regular Measures Theorem and the fact that the product of a finite (or countable) number of \bar{M}_{02} T.M. spaces is \bar{M}_{02} . For example, if $(X_k, \mathcal{U}_k, \mathcal{B}_k)$ are \bar{M}_{02} with p-metrics d_k for $k = 1, \dots, n$, then the p-metric d^n on X^n , defined by

$$d^n(x^n, y^n) = \sum_{k=1}^n d_k(x_k, y_k),$$

generates the product topology \mathcal{U}^n ; furthermore $\mathcal{B}^n = \overline{\mathcal{U}^n}$, and X^n is \bar{M}_{02} .

The Consistency Theorem has many important applications in modern probability theory. For example, Kolmogorov has used it to prove a generalization of Bochner's Theorem on characteristic functions and positive definiteness (compare LOEVE, p. 207, Theorem A and GRENNANDER, p. 134,

Theorem 6.2.4) from the real line to Hilbert space, which in turn is used in the proof of a generalization of the Central Limit Theorem (compare LOEVE, p. 274, Theorem A and GRENANDER, p. 145, Theorem 6.5.1) from real r.v.'s to Hilbert r.v.'s. The Consistency Theorem can also be used to construct pr. distributions on direct sums of Hilbert spaces from given consistent pr. distributions on the finite direct sums (see GRENANDER, p. 139, Theorem 6.2.5), which has applications in the theory of Hilbert-valued stochastic processes.

APPENDIX

LIST OF ABBREVIATIONS AND SYMBOLS

<u>Abbreviations</u>	<u>Meanings</u>
iff	if and only if
\triangleright	implies
top.	topological
nhd.	neighborhood
cont.	continuous
p-metric	pseudo-metric
T.M.	topological measurable
NLS	normed linear space
MCT	Monotone Convergence Theorem
DCT	Dominated Convergence Theorem
pr.	probability
r.v.	random variable
a.s.	almost surely
expt.	expectation
c. expt.	conditional expectation
ind.	independent

<u>Symbols</u>	<u>Meanings</u>
Ω	measurable space with σ -field \mathcal{A} measure space with σ -field \mathcal{A} and measure μ probability space with σ -field \mathcal{A} and probability P
ω	points in Ω
X	measurable space with σ -field \mathcal{B} topological space with topology \mathcal{U} linear space with scalar field Z any combination of the above
x, y	points or vectors in X
X, Y	functions on Ω into X
α, β	scalars (numbers) in Z
f, g	functions on X into Z
$ x $	norm of a vector x
$ \alpha $	absolute value of a scalar α
θ	zero vector

Set Theoretic
Symbols

\emptyset	empty set
$\mathcal{S}(X)$	class of all subsets of X
A^c	complement of a set A
I_A	indicator of a set A
$A + B$	sum of two disjoint sets
$\sum_1^n A_k$	sum of n disjoint sets
$\sum A_n$	countable sum
$\bigcup_{\zeta} C$	$= \bigcup_{C \in \zeta} C$
\mathcal{U}^c	$= [U^c : U \in \mathcal{U}]$

<u>Set Theoretic Symbols</u>	<u>Meanings</u>
\mathcal{B}_C	$= [BC : B \in \mathcal{B}]$
$X^{-1}\mathcal{B}$	$= [X^{-1}B : B \in \mathcal{B}]$
I	index set
X^I	$= \prod_{t \in I} X_t$
X^n	$= \prod_{k=1}^n X_k$
$\tilde{\zeta}$	minimal topology containing the class ζ
$\bar{\zeta}$	minimal σ -field containing the class ζ
\mathcal{F}_{ζ}	minimal field containing the class ζ
$\hat{\zeta}$	minimal monotone class containing the class ζ
ζ^r	minimal ring containing the class ζ
ζ^s	minimal σ -ring containing the class ζ
<u>Topological Symbols</u>	
\mathcal{N}_x	nhd. system of the point x
A°	interior of the set A
A^-	closure of the set A
\mathcal{K}	class of all closed compact sets
G_δ	intersection of a decreasing sequence of open sets
<u>Symbols for Types of Topological Spaces</u>	
T_1	points closed
T_2	Hausdorff
C_1	first countable

Symbols for Types
of Topological
Spaces

Meanings


C_2	second countable
L_1	first Lindelöf
L_2	second Lindelöf
M_1	metric
M_2	separable metric
M_0	pseudo-metric
\bar{M}_{02}	separable, complete pseudo-metric
K_1	sequentially compact
K_2	compact
ℓK	locally compact
σK	σ -compact
G_δ	every closed subset is a G_δ
σKG_δ	σK and G_δ
$\Pi[\sigma KG_\delta]$	product of a finite number of σKG_δ spaces

Measure and pr.
Theoretic Symbols

\mathcal{M}	class of all measurable functions
\mathcal{M}_s	class of all simple functions
\mathcal{M}^+	class of all nonnegative measurable functions
\mathcal{L}	class of all summable functions ($= \mathcal{L}^1$)
\mathcal{L}^2	class of all square summable functions
$\int X d\mu$	integral of the function X
$\int X$	$= \int X d\mu$, where μ is understood

Measure and pr.
Theoretic Symbols

Meanings

EX	$= \int X dP$, expectation of a r.v. X
σ^2_X	$= E X - EX ^2$, variance of a r.v. X
$\ X\ $	\mathcal{L} or \mathcal{L}^2 norm of a Banach r.v. X
$X : Y$	$= \int X \cdot Y$, dot product of two Hilbert r.v.'s, $X, Y \in \mathcal{L}^2$
	end of proof.

BIBLIOGRAPHY

- AKHIESER, N. I. [1], Theory of Approximation, New York: Ungar, 1956.
- AKHIESER, N. I. [2] and Glazman, I. M., Theory of Linear Operators in Hilbert Space, Vol. I, New York: Ungar, 1961.
- BANACH, S., Theorie des Opérations Linéaires, Warsaw, 1932.
- BLACKWELL, D., "On a Class of Probability Spaces," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. II, Berkeley: Univ. of California, 1956.
- BOCHNER, S., "Integration von Funktionen deren Werte die Elemente eines Vektorraumes sind," Fundamenta Math. Vol. 20, 1935.
- BOHNENBLUST, H. F. and Sobczyk, A., "Extension of Functionals on Complex Linear Spaces," Bulletin of the American Math. Soc., Vol. 44, 1938.
- DRIML, M. and HANS, O., "Conditional Expectations for Generalized Random Variables," Transactions of the Second Prague Conference, Prague: Czechoslovak Academy of Sciences, 1960.
- DUBINS, L. E., "Generalized Random Variables," Transactions of the American Math. Soc., Vol. 84-85, 1957.
- DUNFORD, N [1], "Integration in General Analysis," Transactions of the American Math. Soc., Vol. 37, 1935.
- DUNFORD, N. [2] and Schwartz, J. T., Linear Operators, Vol. I, New York: Interscience, 1958.
- GRENNANDER, U., Probabilities on Algebraic Structures, New York: Wiley, 1963.
- HAHN, H., "Über Lineare Gleichungen in Linearen Räumen," Journal für reine und angew. Math., Vol. 157, 1927.
- HALMOS, P. R. [1], Measure Theory, Princeton, N. J.: Van Nostrand, 1950.
- [2] Introduction to Hilbert Space, 2nd ed., New York: Chelsea, 1951.
- HANS, O., "Generalized Random Variables," Transactions of the First Prague Conference, Prague: Czechoslovak Academy of Sciences, 1957.
- HAUSDORFF, F., Set Theory, New York: Chelsea, 1957.

- HILLE, E. and Phillips, R., Functional Analysis and Semi-groups, revised ed., Providence: American Math. Soc., 1957.
- KELLEY, J. L., General Topology, Princeton, N. J.: Van Nostrand, 1955.
- KOLMOGOROV, A. N. [1], Foundations of the Theory of Probability, 2nd English ed., New York: Chelsea, 1950.
- KOLMOGOROV, A. N. [2] and Fomin, S. V., Functional Analysis (2 Vol.), Rochester, N. Y.: Graylock, 1957, 1961.
- LIUSTERNIK, L. and Sobolev, V., Elements of Functional Analysis, New York: Ungar, 1961.
- LOEVE, M., Probability Theory, 3rd ed., Princeton, N. J.: Van Nostrand, 1963.
- LOOMIS, L. H., An Introduction to Abstract Harmonic Analysis, New York: Van Nostrand, 1953.
- McSHANE, E. J. and Botts, T. A., Real Analysis, Princeton, N. J.: Van Nostrand, 1959.
- NAIMARK, M., Normed Rings, Groningen: P. Noordhoff, 1959.
- NEDOMA, J., "Note on Generalized Random Variables," Transactions of the First Prague Conference, Prague: Czechoslovak Academy of Sciences, 1957.
- PARZEN, E., Modern Probability Theory, New York: Wiley, 1960.
- PETTIS, B. J., "On Integration in Vector Spaces," Transactions of the American Math. Soc., Vol. 44, 1938.
- RIESZ, F. and Sz.-Nagy, B., Functional Analysis, New York: Ungar, 1955.
- SEGAL, I. E., "Abstract Probability Spaces and a Theorem of Kolmogorov," American Journal of Math., Vol. 76, 1954.
- SIMMONS, G. F., Introduction to Topology and Modern Analysis, New York; McGraw-Hill, 1963.
- SPACEK, A., "Probability Measures in Infinite Cartesian Products," Illinois Journal of Mathematics, Vol. 4, 1960.
- TAYLOR, A. E., Introduction to Functional Analysis, New York: Wiley, 1958.
- VON NEUMANN, John, Functional Operators, Vol. II, Princeton, N. J.: Princeton Univ., 1950.

ZAANEN, A. C. [1], Linear Analysis, 3rd ed., Amsterdam: North-Holland Publ. Co., 1960.

— [2] Theory of Integration, Amsterdam: North-Holland Publ. Co., 1958.